The Dissertation Committee for Ryan Lee Boyd Certifies that this is the approved version of the following dissertation:

The Multifaceted Measurement of the Individual through Language

Committee:

James W. Pennebaker, Supervisor

Samuel D. Gosling

William B. Swann

Matthew Lease

The Multifaceted Measurement of the Individual through Language

by

Ryan Lee Boyd, B.A.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2017

Dedication

To those very few people who have encouraged my growth through their endless thoughtfulness, kindness, and generosity over the years. I have been more than lucky to receive your support and guidance. I promise to perpetually pay it forward.

The Multifaceted Measurement of the Individual through Language

Ryan L. Boyd, Ph.D.

The University of Texas at Austin, 2017

Supervisor: James W. Pennebaker

Historically, research within the psychological sciences has adopted a classical approach to understanding the individual. This approach regularly involves the observation and measurement of specific, isolated psychological phenomena in an attempt to better understand psychological features, tendencies, and processes at varying levels of interest. While the scope of the traditional approach can vary depending on the construct under investigation, the core methodology and analytic strategy typically adheres to the "isolate and/or manipulate" doctrine for seeking knowledge. In recent years, however, technology has revolutionized researchers' access to computational power, analytic techniques, and even the quality and quantity of data that can be used in scientific pursuits. This dissertation consists of 3 sets of studies that are either a) already published in peer-reviewed journals or b) are currently under review in peer-reviewed journals.

The primary theme to be found in the included studies is a transition from classical methods of assessment to one where the individual is simultaneously quantified in high-dimensional space using language analysis techniques. This approach essentially constitutes the quantification of the individual as a cluster of traits/processes by means of psychological traces that are embedded in (and can be measured indirectly via) a person's language. This approach entails measuring psychological phenomena at both greater depth

and breadth than commonly seen in the psychological sciences and, additionally, serves as a convenient and powerful replacement of traditional approaches to studying psychology in the real world.

The studies included in this dissertation demonstrate the usefulness of a high-dimensional psychometric approach via language in realms of authorship attribution and value measurement. In 2 of the 3 studies, language analytic techniques are used to measure consistencies within the individual that can be capitalized upon in order to determine authorial identities. In the third study, the high-dimensional approach is applied to the realm of values, demonstrating greater utility in a classic research paradigm that vastly outperforms the traditional self-report method.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Historically, research within the psychological sciences has adopted a classical approach to understanding the individual. Under such an approach, researchers regularly gather observations and measurements of specific, isolated psychological phenomena in an attempt to better understand psychological features, tendencies, and processes at varying levels of molecularity. For example, individuals interested in low-level cognitive processes that contribute to broader personality manifestations may conduct experiments using basic cognitive probes (see Robinson, Boyd, & Liu, 2013), whereas psychologists interested in externalized manifestations of personality might observe or measure the behavioral impact of an individual upon their environment (e.g., Graham, Sandy, & Gosling, 2011). While the scope of traditional research methods often vary as a function of the construct under investigation, the core methodology and analytic strategies typically adhere to the "isolate and/or manipulate" doctrine for seeking knowledge (e.g., Goodwin & Goodwin, 2013).

In recent years, however, technology has revolutionized researchers' access to computational power, analytic techniques, and even the quality and quantity of data that can be used in scientific pursuits. Researchers now have unfettered access to vast quantities of real-world data generated by humans in a spontaneous, unprompted manner. While most of this data is in the format of unstructured data (Dell EMC, 2012), new strategies that facilitate the conversion of unstructured data into statistically actionable metrics are rapidly emerging (e.g., Tsai, Lai, Chao, & Vasilakos, 2015). Importantly, one of the most prevalent forms of unstructured data is that of natural language, which has a long history of study within the psychological sciences (see Boyd, in press).

In the past 2 decades, a host of techniques have been developed under continual refinement that impose structure on natural language, such as part-of-speech tagging and

distributed representation modeling (e.g., Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). However, many modern language analysis techniques exist that are explicitly designed for drawing psychological inferences. Modern psychological text analysis techniques differ from those hailing from other traditions primarily in their intended use as well as their psychometric properties when treated as measurement techniques. In *psychological* text analysis methodologies, techniques have been developed that convert raw text into a wide vector of validated measures with established psychometric properties (e.g., Boyd, in press; Boyd & Pennebaker, 2015a). For example, it is possible to concurrently estimate various psychological attributes of the individual such as depression (Stirman & Pennebaker, 2001), personality (e.g., Yarkoni, 2010), political motivations (e.g., Fetterman, Boyd, & Robinson, 2015), and thinking style (Pennebaker, Chung, Frazee, Lavergne, & Beaver, 2014) from a single writing session, social media presence, or verbal exchange.

More broadly, theoretical constructs can themselves be explored using language-based assessment techniques (e.g., Chung & Kramer, 2011; Sagi & Dehghani, 2013). The combination of modern language analysis methodologies with real-world natural language data has, very recently, allowed us to study an individual's psychological properties in a highly refined, ecologically valid manner that far surpasses past efforts in terms of breadth and resolution.

This dissertation consists of 3 sets of studies that are either a) already published in peer-reviewed journals (Boyd & Pennebaker, 2015b; Boyd et al., 2015) or b) are currently under review in peer-reviewed journals (Boyd, under review). As such, the contents of this dissertation should be considered as a general recreation of each of these 3 papers/manuscripts from pre-published documents, and the final works should be referenced directly for a more full consideration of the research conducted. Each of these

studies provides a strong illustration of the strengths of multifaceted psychological assessment via natural language data.

The primary theme to be found in the included studies is a transition from classical methods of assessment to one where the individual is simultaneously quantified in high-dimensional space using language analysis techniques. This approach essentially constitutes the quantification of the individual as a cluster of traits/processes by means of psychological traces that are embedded in (and can be measured indirectly via) a person's language. This approach entails measuring psychological phenomena at both greater depth and breadth than commonly seen in the psychological sciences and, additionally, serves as a convenient and powerful replacement of traditional approaches to studying psychology in the real world.

Chapter 2: Psychological Fingerprinting[1]

INTRODUCTION

In 1728, Lewis Theobald published a play entitled *Double Falsehood*. In presenting this work, he reported that it was based on three original manuscripts of a play that he had discovered, all written by Shakespeare. At the time, Theobald had published extensively on Shakespeare's work and was an avid collector of playwright manuscripts (Corbett, 1744). Unfortunately, Theobald's original manuscripts are believed to have been lost in a library fire (Carnegie & Taylor, 2012). The authorship underlying *Double Falsehood* has now been contested for centuries (Dominik, 1991), with scholars having offered evidence of the play being written either by Shakespeare or Theobald himself (see Brean Hammond's [2010] edited work, *Double Falsehood* for a thoughtful set of analyses).

Double Falsehood is particularly interesting because later scholars found references to a similarly-themed play presented in London in 1613 called *The History of Cardenio* by Shakespeare and John Fletcher. Before his death in 1616, Shakespeare coauthored at least two other plays with Fletcher, *Henry VIII* and *Two Noble Kinsmen*. In the current research, we introduce new techniques combining contemporary authorship identification methods with the psychology of language to infer the writer, or writers,

[1] Citation for the published version of this chapter: Boyd, R. L., & Pennebaker, J. W. (2015). Did Shakespeare write Double Falsehood? Identifying an individual's psychological signature with text analysis. *Psychological Science, 26*(5), 570-582. The author of this dissertation (Ryan L. Boyd) was the primary researcher for this study and was the principle individual involved in the data analyses and writing.

behind *Double Falsehood* based upon its high-dimensional "psychological signature."

Authorship Identification: A Brief Primer

Historically, many methods of authorship identification (AID) have existed. Perhaps the most well-known is "stylometry" (Holmes, 1994). Traditional stylometry assumes that language patterns are acquired idiosyncratically, resulting in each person's unique use of words (Van Halteren et al., 2005). Early stylometry examined basic language features including spelling (Craig, 1999; Wellman, 1936), vocabulary (Ule, 1982; Johnson, 1996), complexity (Fucks, 1952; Morton, 1978), and the physical properties of documents (e.g., Lerner & Lerner, 2005). Viewed as clues about authors' personalities, cultures, and experiences, these variables shaped hypotheses about an unknown author's identity. Importantly, however, these methods were of limited usefulness when considered individually (Grieve, 2007).

The first computer-based stylometric analysis was applied to eleven of the 84 Federalist Papers by Mostellar and Wallace (1964). The authorship of the eleven papers was disputed and, by comparing their use of function words, Mostellar and Wallace concluded that all were written by one particular author who had written the majority of the other Federalist papers (see Juola, 2008, and Pennebaker, 2011). Despite the apparent success of the computer-based approach, AID methods often provide only probabilistic clues as to a document's authorship.

The Psychology of Language

While dozens of quantification methods exist, those linked to AID often focus on patterns of individual words (i.e., "unigrams") and phrases (e.g., "bigrams", "trigrams", etc.; see Koppel et al., 2008). However, words can be classified along hundreds of *psychological* dimensions as well, including cohesion, time orientation, and sentiment, to name a few. Perhaps the most basic distinction among words from a psychological perspective is between content and function words (e.g., Miller, 1995). Function words include conjunctions, prepositions, and related words. In the English language, there are relatively few common function words, yet they account for the majority of written/spoken words (Pennebaker, Mehl, & Niederhoffer, 2003). Recent studies find that function words reveal much about psychological and social processes, including emotional state (Stirman & Pennebaker, 2001), cognitive complexity (e.g., Bond & Lee, 2005), and sociability (e.g., Beukeboom, Tanis, & Vermeulen, 2013). See Tausczik and Pennebaker (2010) and Chung and Pennebaker (2007) for detailed links between function words and social/psychological processes.

On the other hand, most of the English vocabulary is content words. Content words reveal psychological information in a more transparent way than function words, conveying the "who", "what", "when", etc., of mental life. For example, people may convey negative emotions with words of anger or anxiety (e.g., Back, Kufner, & Egloff, 2010; Pennebaker, 2004). Securely attached people tend to use more words related to inclusion (Cassidy, Sherman, & Jones, 2012). A message's content can indicate a person's culture and time (e.g., Leigh, 2011), and socially-connected people make more

social references in their self-concepts (Burke & Dollinger, 2005). Even a person's preoccupations (e.g., food, drinking, sex) are often apparent in the content of one's language and predictive of later behaviors (Robinson, Navea, & Ickes, 2013).

Importantly, language use is consistent within person across time and context (Mosteller & Wallace, 1964; Pennebaker & King, 1999; Pennebaker, 2011). We highlight two primary implications of this: 1) a person can be mapped onto multiple psychological dimensions simultaneously via their unique language, and 2) this psychological mapping will be relatively stable across time and context for most individuals.

In sum, function and content words simultaneously reflect many different psychological patterns and processes unique to the individual.

Current Study: Creating a Psychological Signature of the Individual

In the current study, we bridge the gap between the field of AID and current psychological understandings of language by introducing the concept of language-derived "psychological signatures" to differentiate individuals. At the heart of modern AID is a loosely-grouped assortment of procedures known as "machine learning" (see: Koppel, Schler, & Argamon, 2008). *Machine learning* refers to techniques whereby computers are taught to discriminate between outcomes or categories. For example, a computer can be trained to make difficult medical diagnoses based on what symptoms are (and, importantly, are not) present (Kononenko, 2001). These techniques can even identify faces and objects with unsettling accuracy (Viola & Jones, 2004). Such

discriminative power has obvious appeal to those seeking to solve questions of authorship.

While machine learning procedures use multiple measures to demarcate outcomes, they have not been applied to explicitly psychological interpretations of language. By considering multiple psychological dimensions simultaneously, we are able to create a "psychological signature" of a person derived entirely from their language. This high-dimensional composite of an individual's mental life represents the dimensions along which someone thinks, feels, and engages with the world in a way that is uniquely their own. Importantly, these representations of people's mental worlds not only differentiate individuals, but provide powerful clues as to how they differ from one another in specific and fundamental psychological terms.

The current project applies new language analytic strategies to the curious case of *Double Falsehood*, a play of disputed origins, by introducing the concept of psychological signatures. Four new and one traditional method of AID are described. These methods are then used in the course of modern classification procedures to explore how they compare to regions of the psychological signature fashioned from *Double Falsehood*. Results are discussed in terms of their interpretation, convergence with observation/life outcome data, and implications.

METHODS

The three most likely authors of *Double Falsehood* have been proposed by previous scholars to be William Shakespeare, John Fletcher, and Lewis Theobald (e.g.,

Hammond, 2010). Texts from each author were acquired from various sources, resulting in a total of 55 texts for analyses: 33 plays by Shakespeare, 9 by Fletcher, 12 by Theobald, and *Double Falsehood* – only plays that are generally believed to have been written in solo by each author were used (see Appendix for a list of plays included; see also: Boyd & Pennebaker, 2015b). Each text was manually stripped of extraneous information that did not directly reflect the author's language; this included text such as its publication information (e.g., title, author name), the list of *dramatis personae*, and appendices. Stage directions were left intact.

All cleaned texts were processed through software designed specifically to convert idiosyncratic and outmoded spellings to their United States equivalents (e.g., "threat'ning" to "threatening", "prithee" to "pray thee"; for a complete list, see Boyd, 2014c). While the conversion process was by no means exhaustive, it improved analytic reliability, both within and between authors. Copies of all modified text files are available from the authors.

Language Quantification

The quantification techniques used in the current research constitute what is referred to as a "word counting" approach. With this approach, indices of language are expressed as a percentage of the language category's prominence relative to the whole document. A word counting approach to language assessment can seem superficial on many levels. First, it ignores context – the same words often have different meanings in different situations (Nguyen & Ock, 2013). A person can say the same sentence in a

genuine, ironic, or sarcastic manner, thereby conveying completely different word meanings. Second, word counting is viewed by some as being prone to error in that the same word can inherently have different meanings (e.g., Schwartz et al., 2013). The word "depressed", for example, can variously refer to sadness, economics, or even the physical state of an object.

These are valid criticisms when considering single words or sentences. Our approach, however, takes a broader view and adopts a probabilistic model (e.g., Harris, 1954). Statistically speaking, the majority of times where people use the word "depressed", they are referring to the psychological condition (Savova et al., 2007). Moreover, a truly melancholic person will tend to use a variety of other depression-related words (Ramirez-Esparza, Chung, Kacewicz, & Pennebaker, 2008). The current techniques capitalize on these human tendencies by counting the proportion of all category-relevant words within a text, be it a political speech, piece of literature, transcribed conversation, or Facebook post. For any given text, it is possible to use a text analysis program such as Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007) to calculate the percentage of each language dimension based on the total words in the text; this is the chief quantification procedure used for the current sample.

Function word measures

High-frequency function words are commonly used as variables in AID studies (see Koppel et al., 2008). However, there are eight overarching classes of function words:

personal pronouns, impersonal pronouns (e.g., it, any, thing), articles, prepositions, auxiliary verbs, conjunctions, negations, and high frequency adverbs lacking direct referents (e.g., very, really, so). These eight classes of function words have been shown to reflect separate psychological processes through their use (Tausczik and Pennebaker, 2010). On average, the total percentage of function words in non-technical texts average around 52-60% (Pennebaker et al., 2007); in the current sample, the average rate of function word use was 53.4%. By quantifying the use of function word classes in *Double Falsehood*, it is possible to explore the probability that Shakespeare, Fletcher, and Theobald contributed to the thinking style of the play. Additionally, this procedure allows one to identify the unique psychological characteristics typical of each of the three candidate authors.

The Categorical-Dynamic Index and Complexity

Statistical analyses of function words across all types of text typically reveal a single dimension of language use that is called the categorical-dynamic index (CDI; Pennebaker et al., 2014). The CDI is a continuum along which any text necessarily falls. At the categorical end of the continuum, people tend to use high rates of nouns, articles, and prepositions. A closer inspection of categorical texts finds that people who are high on this dimension tend to be more analytic or formal in their thinking. This means that they tend to classify objects, people, and events in hierarchical ways – people high in categorical thinking tend to be more emotionally distant and problem-solving in their approaches to everyday situations.

At the other end of the CDI dimension are texts with high rates of auxiliary verbs, pronouns, adverbs, and the other function word categories. People who are dynamic thinkers tend to live more in the here-and-now, tell stories, and are more focused on social matters (see Pennebaker, 2011, for more details and an overview of the distinction between categorical versus dynamic thinking).

In the current research, "categorical complexity" is considered alongside two other conventional forms of complexity that have been used in past AID and psychological research: average sentence length (e.g., Yule, 1944) and the use of large words (e.g., Brinegar, 1963). As with function word composites like the CDI, the use of both longer sentences (i.e., more words per sentence) and of large words are psychologically meaningful from a cognitive perspective (e.g., Guastella & Dadds, 2006; Hartley, Pennebaker, & Fox, 2003) and are quite reliable (Pennebaker & King, 1999). Additional discussion of the CDI is presented in Boyd and Pennebaker (2015b).

Content Words

The analysis of content words was accomplished in two ways. The first was to rely on the default content categories used in the computer program LIWC. The second, which is described in the next section, was to rely on a meaning extraction technique to inductively identify themes within the plays.

In the same way that a text can be scanned for the eight classes of function words, they can be scanned for well-established categories of content words. The default LIWC 2007 dictionary (Pennebaker et al., 2007) codes for words that belong to over 40 content

categories, including words related to positive and negative emotions, family members, sensory perceptions, religion, death, etc. This dictionary is the most widely used in psychology (Schwartz et al., 2013b) and its psychometric properties have been extensively validated across time, location, and even multiple languages (e.g., Markus et al., 2008; Pennebaker et al., 2007).

The Meaning Extraction Method

Within the last decade, a number of computerized methods have been developed that allow researchers to automatically extract "themes" from large bodies of text. One such technique from the field of psychology is called the Meaning Extraction Method, or MEM (Chung & Pennebaker, 2008). The MEM procedure was applied to all plays using specialized software (see Boyd, 2014a), resulting in 13 broad themes. The presence of these themes can then be measured in a sample of text by counting the number of words for each theme relative to the text as a whole – this was done for *Double Falsehood* and the works of Shakespeare, Fletcher, and Theobald using word counting software (Boyd, 2014b).

Importantly, for any single theme, two or more of the plays may be virtually indistinguishable on average – for example, all authors (generally speaking) use very similar, relatively low rates of words contained in a broader "family structure" theme. However, when looking at an author in total, something akin to a "signature" begins to arise. One can then combine how each author tends to weave together all of the themes across all of their works – this results in a more general "thematic signature" for each

author. Just as one's personal signature tends to be a little different every time it is written and depending on the type of document signed (e.g., a birthday card versus a business document), a thematic signature will exhibit variations from piece to piece. As with other content categories, the categories of a thematic signature exhibit reliability across time. The relative presence or absence of a given theme can be a useful cue to an individual's psychological characteristics (Chung & Pennebaker, 2008; Lowe et al., 2013; Ramirez-Esparza et al., 2008). Boyd and Pennebaker (2015b) contains additional discussion, information, and statistics relevant to the MEM.

Traditional stylometric measure: Linguistic "tells" with low base rate words

In high-stakes poker, experts frequently analyze the ways in which their opponents laugh, speak, or fidget when they are concealing a particularly good (or very bad) hand. The belief is that people have subtle "tells" – specific behaviors that reveal their anxiety or excitement. A similar idea has been used by stylometrists and other language experts in author identification. Specifically, many writers tend to use idiosyncratic words and phrases across multiple writings. Just as various authors used different function words at different rates, people also differ in their use of relatively uncommon words and phrases (e.g., Craig & Kinney, 2009; Vickers, 2011). For example, Foster (1996) was able to identify the anonymous author of *Primary Colors* as Joe Klein due, in part, because of the way Klein consistently used relatively obscure adverbs (e.g., goofily, handily, huffily, juicily) in his newspaper and magazine articles as well as in *Primary Colors*. Traditional stylometry, then, identifies words used at a low rate in the

general population but consistently by selected authors across a series of published works.

The psychology of low base rate tell words is rather different from the analysis of language style or content. As mentioned earlier, *how* people speak and *what* they speak about reveal basic psychological tendencies about their thinking, perceptions of the world, and connection with others. Use of low base rate words, on the other hand, are less psychologically meaningful. Often, they likely reflect chance experiences in the people's lives that pertain to language learning in the family and school or their personal aesthetics (e.g., Colman, Walley, & Sluckin, 2011). Shakespeare might have used the word *behalf* simply because he liked the sound of it. Fletcher may have sprinkled the word *handsomely* into most of his plays because an admired protégé used the word. Although not likely to be deeply psychological, low base rate words can be clues to authorship because they consistently emerge from each of us. Additional information on our identification of low base rate "tell" words is located in Boyd and Pennebaker (2015b).

RESULTS

While there are dozens of classification procedures that fall under the umbrella of supervised machine learning and have been used in AID (see Juola, 2008, and Koppel et al., 2008, for reviews), we focus on three: linear discriminant analysis (LDA), decision trees (DT), and support vector machines (SVM). Additional discussion of all three methods are presented in Boyd and Pennebaker (2015b). Descriptive statistics for all

measures, and a naïve conceptualization of distance for those unfamiliar with our statistical approaches, are presented in Boyd and Pennebaker (2015b).

LDA, DTs, and SVMs behave in very different ways in both mathematical and procedural terms, yet may be viewed as complementary to one another in practice (e.g., Curram & Mingers, 1994; Chang, Guo, Lin, & Lu, 2010). As such, the current research separately employs these procedures to look for convergence. Just as language variables can lack predictive strength in isolation, so too are complementary classification methods able to provide better information when considered together (e.g., Kacmarcik & Gamon, 2006; Martindale & McKenzie, 1995; Somers, 1998). For all analyses, we used Fisher's (1936) classical LDA (with equal prior probabilities assigned), the J48 DT algorithm (known for its power and relative simplicity; Witten, Frank, & Hall, 2011), and the sequential minimal optimization support vector machine (SMO SVM; Hall et al., 2009) to infer the authorship probabilities in *Double Falsehood*.

Whole Play Analyses

All plays were quantified with the five previously-described techniques, then were classified, by author, using LDA, J48 DTs, and SMO SVMs. *Double Falsehood* was then allowed to be freely classified and assigned to any one of the three authors (Shakespeare, Fletcher, and Theobald) by these three analytic strategies. The main results are presented in Table 2.1. Primary statistical reports and cross validation details are presented in Boyd and Pennebaker (2015b).

	LDA	*p*	J48 DT	*p*	SMO SVM	*p*
Function Word Classes	Shakespeare	91.4%	Shakespeare	96.8%	Shakespeare	83.3%
CDI / WPS / Large Words	Shakespeare	61.0%	Shakespeare	93.1%	Shakespeare	78.9%
LIWC Content Categories	Theobald	97.3%	Shakespeare	97.1%	Shakespeare	75.4%
Thematic Signatures	Shakespeare	100%	Shakespeare	97.1%	Shakespeare	99.8%
Low Base Rate "Tells"	Shakespeare	83.8%	Shakespeare	100%	Shakespeare	97.1%

Table 2.1: Results for each language measure, by classification technique.

Note: Authorship likelihood estimates are presented as “best candidate” probabilities for LDAs and prediction margins for the J48 decision trees and SMO SVMs.

Function Word Results

The 8 classes of function words

All three models designated Shakespeare as the best authorial candidate for *Double Falsehood* when considering the 8 general classes of function words. Generally speaking, LDA and SMO SVM approaches were able to discriminate between the authors using vectors comprised of all 8 classes of function words (see Figure 2.1, left side). Theobald was primarily distinguishable by his high use of prepositions and articles and low use of other stylistic categories of language, whereas Fletcher was quite the opposite. Shakespeare was able to be differentiated as stylistically trending towards Fletcher, but moderately enough to be distinct. The J48 DT came to the same conclusion (i.e., Fletcher being stylistically high in more “dynamic” language variables, such as auxiliary verbs and adverbs), but required fewer of the 8 classes to successfully discriminate between authors. These results show that, indeed, all three authors have distinct stylistic

psychological signatures along function word dimensions, and that *Double Falsehood*'s stylistic composition is, on the whole, most analogous to that of Shakespeare.

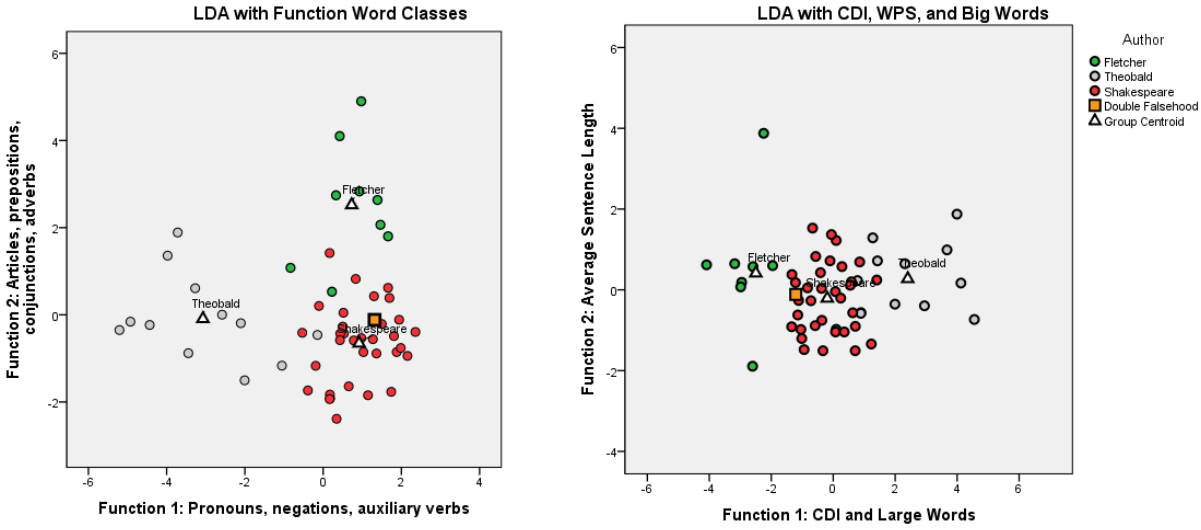

Figure 2.1: Results for the LDAs using the 8 classes of function words (left) and cognitive / stylistic complexity measures (right).

CDI and complexity

As with the 8 classes of function words, all three models assigned *Double Falsehood* to Shakespeare with a high probability, and analyses found unique degrees of cognitive complexity for the three authors. Analyses found Theobald to be the most complex in terms of CDI and both conventional complexity measures (average WPS and use of large words). At the other other end of the spectrum, Fletcher exhibited the most dynamism along the CDI, but was somewhat higher in conventional complexity. As with the 8 classes of function words, Shakespeare was somewhere in the middle, but trending

towards Fletcher's levels of conventional complexity (relative to Theobald; see Figure 2.1, right side). Similarly, the DT found that CDI was the most robust discriminating metric, and relied only on this measure to distinguish between the three authors.

Content Words Results

LIWC Content Categories

The LDA, DT, and SMO SVM models were able to successfully discriminate between authors when considering the LIWC content categories that were typical of each author. LDA created 2 vectors that were largely (but not entirely) composed of two classes and subclasses of contents words: 1) words related to emotion, both positive and negative, and 2) words related to thought processes (i.e., cognitive mechanisms) and social processes. The DT used similar categories to discriminate between authors, but only relied on the broadest emotion category (labeled "affect") and a specific subtype of cognitive mechanism ("certainty"). Generally speaking, Theobald scored high in emotional words and lowest in cognitive mechanism words, whereas Fletcher showed a reversal of this pattern Shakespeare's content exhibit the lowest levels of the emotional vector by far, but scored similar to Theobald in the cognitive mechanism vector score.

In this analysis, LDA disagreed with the DT and SMO SVM on the most likely authorial candidate, with the former asserting Theobald's content fingerprint being dominant in *Double Falsehood* as a whole and the latter two indicating Shakespeare (see Figure 2.2, left side). In total, then, there appears to be consensus of Shakespeare's primary influence, albeit non-unanimous.

Meaning Extraction

All three classification procedures were able to discriminate between authors based on their thematic signatures using the 13 themes extracted using the MEM. LDA used two vectors generally composed of 1) high use of the "Nobility" and "Femininity" themes and low rates of "Emotionality" and "Romance" themes, and 2) higher levels of the "Social", "Youth", and "War and Battle" themes, with low rates of the "Royalty" and "Slumber" themes (see Figure 2.2, right side). Again, the DT only required 2 of these content categories to distinguish authors: 1) the "Emotionality" theme (highest for Theobald) and 2) the "Social" theme (highest for Fletcher). The SMO SVM converged with the other procedures in determining that the thematic composition of *Double Falsehood* most closely resembled that of the other works of Shakespeare with relatively high certainty.

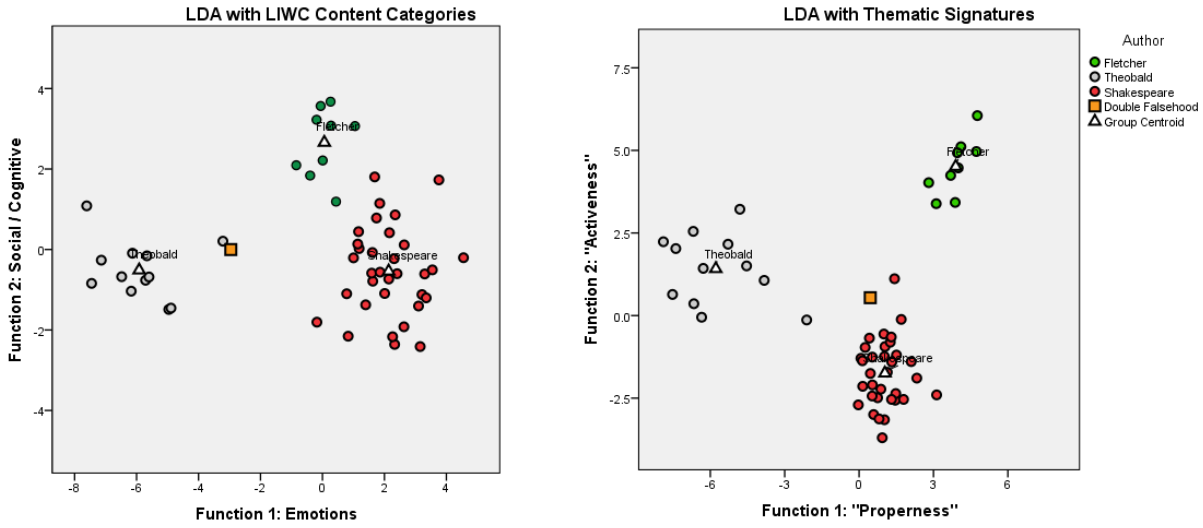

Figure 2.2: Results for the LDAs using the LIWC content categories (left) and MEM thematic signatures (right).

Note: Regarding thematic signatures: Function 1 ("Properness") is largely (but not entirely) a composite of high amounts of the "Nobility" and "Femininity" themes and low amounts of "Emotionality" and "Romance" themes. Function 2 ("Activeness") is largely a composite of the "Social", "Youth", and "War and Battle" themes, with low amounts of the "Royalty" and "Slumber" themes.

Low base rate words.

The LDA, DT, and SMO SVM were able to distinguish between the three authors using their low base rate "tell" words and phrases with considerable ease; this is expected, as these N-grams were specifically selected due to their differentiating properties. Generally speaking, all discriminative procedures relied on a similar strategy: classify plays using 1) higher amounts of Shakespeare's trademark phrases and 2) lower amounts of Theobald's trademark phrases – remaining plays were designated as Fletcher (see Figure 2.3). While *Double Falsehood* contained trademark N-grams that could be

reflective of all three authors, all procedures agreed in assigning *Double Falsehood* to Shakespeare with high likelihood, as his low base rate words and phrases were the most dominant in the play. Boyd and Pennebaker (2015b) presents an alternative AID approach to low base rate "tell" words that offers a different, but similar, viewpoint. Boyd and Pennebaker (2015b) offers a different traditional AID test that relies on function (rather than content) word distributions.

Figure 2.3: LDA of low base rate "tell" words.

ANALYSES BY ACT, METHODS

Thus far, we have used five separate language quantification methods paired with three powerful classification techniques to determine the similarity between each author's psychological signature and that of *Double Falsehood* as whole. Overall, the entire play is consistently linked to Shakespeare with a high probability, making it unlikely that it was forged by Theobald. Knowing that *Cardenio* is generally agreed upon as being the work of both Shakespeare and Fletcher (e.g., Freehafer, 1969; Kukowski, 1990), one interpretation is that they indeed collaborated, with Fletcher bringing a rather small contribution to the table while Shakespeare played the role of "master architect". The clearest finding thus far is the likely role of Theobald. Although we do not see a *total* absence of similarity to Theobald in *Double Falsehood* as a whole, we simply find *very little* of it in probabilistic terms with the exception of one outlying statistic. If the play is a genuine Shakespeare—Fletcher creation, it seems reasonable that any later editing of content may have been made by Theobald himself; this is a behavior in which he commonly engaged (see Carnegie, 2012) and would account for his strong showing in the LDA with LIWC content. However, we caution against deep interpretation, as the other two models did not suggest the same result.

An important criticism of the research strategy thus far is that it is relatively broad and crude. That is, the methods have been designed to look at *Double Falsehood* as a single entity without regard to its constituent pieces. One of the advantages of using machine learning procedures is that it is possible to conduct more nuanced analyses. With nuance, however, comes less certainty. From a statistical perspective, there will be more

variability or "jitter" inherent to our findings as we analyze progressively smaller groups of words. That is, standard word distributions tend to become progressively less reliable as language sample sizes decrease (e.g., Landauer & Dumais, 1997). Nevertheless, each act of *Double Falsehood* is of a sufficient size for us take a closer look at this level.

To gain a better understanding of the authorship patterns of *Double Falsehood*, we separated the play into its five constituent acts (following the cleaning procedures and analyses described earlier). Similarly, all other comparison plays by Theobald, Fletcher, and Shakespeare were automatically segmented into 5 equal pieces, resulting in 275 total observations (55 plays × 5 segments). The comparison plays were segmented to compare the acts of *Double Falsehood* to comparable *samples* of plays, rather than determine how similar each act was to *entire* plays by the three candidate authors. The words from each act were then submitted to the same language quantification and statistical techniques that were described above.

ANALYSES BY ACT, RESULTS

LDA, DT, and SMO SVM procedures were performed that are analogous to those performed on whole plays. The main results of these analyses are presented in Table 2.2. Primary statistical details are presented in Boyd and Pennebaker (2015b). Note that, given the larger number of comparison observations due to segmentation, all three analytic approaches are necessarily somewhat more complicated than before due to greater variability between observations and more language required for precise classification. Nevertheless, most analyses relied on highly similar (and sometimes the

exact same) combinations of psychological language metrics to differentiate authors. Results generally show a strong presence of Shakespeare in the early parts of the play, with Fletcher appearing to make his greatest contributions in the final two acts. Theobald's small presence is seen only in the context of content word analyses and low base rate "tell" words. See Boyd and Pennebaker (2015b) for additional discussions and visualized results for all "by act" analyses.

DISCUSSION

Was *Double Falsehood* written by Shakespeare and Fletcher, or is it a well-executed forgery by a man who was knowledgeable in both theatre practice and Shakespeare's many works? Across analyses of style, content, and low base rate words, a consistent psychological signature emerges that is consistent with the writings of Shakespeare and Fletcher. Moreover, the results involving the five acts of *Double Falsehood* overlap with much of the general scholarly consensus. Like others, we find a markedly Fletcherian trend in the final acts of the play (Folkenflik, 2012). Other inquiries have found hints of separate stylistic and content contributions of Shakespeare and Fletcher throughout the play, as does ours (cf., Stern, 2011).

Notable is the general absence of Theobald's psychological signature. However, he does make passing statistical appearances in terms of content words, which are more subject to intentional insertion than are function words. Importantly, Theobald is well-known not only for *Double Falsehood* but his editorial choices as well, making it unlikely that he would leave *Cardenio* wholly untouched (Carnegie, 2012). Nevertheless,

our results offer consistent evidence against the notion of *Double Falsehood* being Theobald's "whole-cloth forgery" (Dolan, 2013).

LDA Results, by Act	Act I	*p*	Act II	*p*	Act III	*p*	Act IV	*p*	Act V	*p*
Function Word Classes	Shakespeare	95.6%	Shakespeare	88.7%	Fletcher	54.6%	Fletcher	71.4%	Fletcher	82.3%
CDI / WPS / Large Words	Shakespeare	66.6%	Shakespeare	74.1%	Shakespeare	50.8%	Fletcher	64.4%	Fletcher	63.3%
LIWC Content Categories	Shakespeare	99.8%	Theobald	93.2%	Shakespeare	99.7%	Fletcher	99.4%	Shakespeare	71.6%
Thematic Signatures	Shakespeare	99.7%	Shakespeare	96.2%	Shakespeare	90.5%	Fletcher	72.7%	Fletcher	83.1%
Low Base Rate "Tells"	Shakespeare	43.5%	Shakespeare	50.1%	Shakespeare	62.1%	Theobald	56.6%	Shakespeare	45.7%
J48 DT Results, by Act	Act I	*p*	Act II	*p*	Act III	*p*	Act IV	*p*	Act V	*p*
Function Word Classes	Shakespeare	87.2%	Shakespeare	87.2%	Shakespeare	54.5%	Shakespeare	54.5%	Fletcher	96.0%
CDI / WPS / Large Words	Shakespeare	69.3%	Shakespeare	69.3%	Shakespeare	69.3%	Shakespeare	69.3%	Fletcher	83.8%
LIWC Content Categories	Shakespeare	93.4%	Fletcher	61.9%	Shakespeare	93.4%	Fletcher	61.9%	Fletcher	95.0%
Thematic Signatures	Shakespeare	92.9%	Shakespeare	92.9%	Shakespeare	92.9%	Shakespeare	92.9%	Shakespeare	92.9%
Low Base Rate "Tells"	Shakespeare	99.4%	Shakespeare	99.4%	Theobald	100%	Fletcher	87.2%	Shakespeare	99.4%

SMO SVM Results, by Act	Act I	*p*	Act II	*p*	Act III	*p*	Act IV	*p*	Act V	*p*
Function Word Classes	Shakespeare	98.5%	Shakespeare	97.3%	Shakespeare	67.0%	Shakespeare	50.6%	Fletcher	58.5%
CDI / WPS / Large Words	Shakespeare	81.2%	Shakespeare	79.7%	Shakespeare	75.6%	Shakespeare	57.2%	Shakespeare	61.1%
LIWC Content Categories	Shakespeare	95.2%	Theobald	99.6%	Shakespeare	96.3%	Fletcher	97.3%	Fletcher	69.7%
Thematic Signatures	Shakespeare	93.5%	Shakespeare	68.7%	Shakespeare	93.4%	Fletcher	83.5%	Fletcher	83.5%
Low Base Rate "Tells"	Shakespeare	88.4%	Shakespeare	97.4%	Theobald	45.2%	Fletcher	38.9%	Shakespeare	99.6%

Table 2.2: Results for each language measure, by act, by classification technique.
Note: Authorship likelihood estimates are presented as “best candidate” probabilities for LDAs and prediction margins for the J48 decision trees and SMO SVMs.

Psychological signatures and convergence with historical reports.

A promising aspect of our analyses is that the methods allow for the inference of Shakespeare, Fletcher, and Theobald's unique psychological signatures. Recall that function and content words offer different types of psychological information. With regards to both broader categories of language, analyses identify unique thinking styles, as well as thought contents, for all three authors. While the content of thought may be mimicked in a document with some accuracy, psychological dimensions revealed by function words are nearly impossible to forge without computer assistance. The following analysis of the each author's psychological signature is speculative, however, the current results converge with the general scholarly consensus of the three men's historical profiles as well as observer reports, life outcomes, and recorded behaviors.

Perhaps the strongest comparison can be drawn between Fletcher and Theobald who consistently showed opposite patterns of language use. Recent studies suggest that people who use pronouns and auxiliary verbs at high rates tend to be more socially engaged, and enjoy telling stories more than people who use these parts of speech at lower rates. Those who tend to use articles and prepositions at high rates (consistent with high CDI scores) tend to be more organized, logical, and formal in their daily lives (Pennebaker & King, 1999; Pennebaker, 2011).

Recall that Fletcher used more dynamic language than Theobald, who consistently relied on more categorical language. Additionally, Fletcher used relatively high amounts of social LIWC content words and a socially-oriented MEM theme. While

few concrete details of Fletcher's life survive (Shakespeare, Irving, Marshall, & Dowden, 1890), it is known that he had many close, long-lasting personal and professional relationships (Ide, 2010) and is said to have fondly swapped attire with close colleagues (Clark & Aubrey, 1898). There is no evidence that he was particularly scholarly or even particularly organized.

Theobald, on the other hand, left an impressive paper trail suggesting that he was relatively distant and aloof in social terms, yet meticulous and organized elsewhere. Indeed, it is known that he went to great lengths to institute high accuracy in his editorial work (Seary, 1990). Furthermore, Theobald often openly insulted his contemporaries in the process of correcting their mistakes, drawing public recriminations and contempt in the process (e.g., Jones, 1919; Rogers, 2004).

Most of Shakespeare's personal life is shrouded in mystery, and public records are the basis for most assumptions about his life and livelihood (Potter, 2012). However, Shakespeare's psychological signature suggests that he possessed some similarities to both Fletcher and Theobald. Like Theobald, his high use of prepositions suggests an education focusing on grammar (Tausczik & Pennebaker, 2010). Accordingly, most scholars believe that Shakespeare was classically trained in grammar school during his youth (e.g., Barkan, 2001). With regards to social interests, however, Shakespeare appears more similar to Fletcher, with a relatively dynamic writing style and relatively high use of social content words. Again, scholars suggest that Shakespeare was socially focused and interested in climbing higher on the social ladder (see Potter, 2012).

Three important caveats must be considered in interpreting these findings. First, all statistical tests reported here are premised on the belief that only Theobald, Fletcher, and Shakespeare are possible contenders as authors. With more candidates, the probability estimates for "most likely" author would likely shrink. Second, the analyses were broad in nature and did not analyze the plays in a scene-by-scene fashion; closer language analyses may better pinpoint specific contributions. Finally, we assume a fairly informal collaborative style between Shakespeare and Fletcher. In other work, close collaborations have resulted in a writing style that bears weak resemblance to either author. For example, the writing style of jointly-written Lennon / McCartney songs were different from songs that either wrote alone (Petrie, Pennebaker, & Sivertsen, 2008). If *Double Falsehood* was authored by multiple parties as the current research suggests, it is not entirely clear whether each author's distinct psychological signature would be discernible, or at what level of granularity (see Boyd & Pennebaker, 2015b for further discussion).

Conclusion.

In combining the various dimensions of a person's mental life, we are able to not only differentiate individuals, but paint a very rich picture of who they are, how they think, and what they think about via the creation of psychological signatures. Such techniques show promise for authorship identification, but may be extremely valuable for better understanding the individual's composite mental life across multiple disciplines in the psychological sciences.

Chapter 3: Measuring Core Values via Natural Language[2]

INTRODUCTION

The increasing amount of publicly available web data has provided a new lens through which we can study how people are thinking, behaving, and feeling (Lazer et al. 2009). Recent developments in natural language processing and information retrieval techniques have allowed researchers to better understand and model social and psychological processes such as personality (Yarkoni 2010), emotion (Strapparava and Mihalcea 2008), and online behaviors (Zhang et al. 2011). We can now study psychological traits and their links to behaviors on a larger scale than ever before through the analysis of social media data.

The current research explores the psychological construct of values, their measurement, and their relationship with behaviors. Using natural language processing techniques, we analyze the ways in which people describe their personal values and behaviors, then compare them with closed (i.e., "forced choice") self-reports. We then expand our study of how values and behaviors are revealed in language to a large corpus of Facebook status updates. This project raises a central question: How should we measure values? That is, are values best measured through traditional self-reports or can we better assess them through the analysis of natural language? Finally, how are values – as measured either through questionnaires or language – related to behaviors?

[2] Citation for the published version of this chapter: Boyd, R. L., Wilson, S. R., Pennebaker, J. W., Kosinski, M., Stillwell, D. J., & Mihalcea, R. (2015). Values in words: Using language to evaluate and understand personal values. *Proceedings of the Ninth International AAAI Conference on Web and Social Media*, 31-40. The author of this dissertation (Ryan L. Boyd) was the primary researcher for this study and was the principle individual involved in the data analyses and writing.

Values and Value Research

Psychologists, historians, and other social scientists have long argued that people's basic values predict their behaviors (Ball-Rokeach, Rokeach, and Grube 1984; Rokeach 1968). Further, human values are thought to generalize across broad swaths of time and culture (Schwartz 1992) and are deeply embedded in the language that people use on a day-to-day basis (Chung and Pennebaker 2014; Lepley 1957).

In psychological research, the term value is typically defined as a network of ideas that a person views to be desirable and important (Rokeach 1973). Values are usually thought of as relatively abstract, giving rise to a broad constellation of related attitudes and behaviors. For example, a person who values "honesty" will typically hold a very negative attitude towards dishonest politicians and, accordingly, will be less likely to vote for them in the future (for a discussion of the links between values and attitudes, see Kristiansen and Zanna, 1988). Such core values are pervasive and often internalized at a very young age (Aronson 2004). It is generally believed that the values which people hold tend to be reliable indicators of how they will actually think and act in value-relevant situations (Rohan 2000).

Over the years, many researchers have conceptualized different frameworks that are believed to cover nearly all core human values (Rokeach 1968). One of the most accepted and widely used of these frameworks was developed by Schwartz and colleagues around two decades ago (Schwartz et al. 2012). The most prevalent form of this approach to the study of values suggests that there are ten primary values organized into a circumplex. This circumplex serves as the umbrella under which the majority of human value judgments fall. These 10 value types are as follows: self-direction (S-D), stimulation (Stim), hedonism (Hed), achievement (Achiev), power (Pow), security (Sec), conformity (Conf), tradition (Trad), benevolence (Benev), and universalism (Univ).

Schwartz's 10-value model has seen great success in psychological research as well as other fields. The basic circumplex model has been applied to the understanding of culture (Schwartz 1994; 2004), religion (Schwartz and Huismans 1995), cognitive development (Bubeck and Bilsky 2004), and politically-motivated behaviors (Caprara et al. 2006), to name but a few domains. Generally speaking, the vast majority of this research has been built upon the Schwartz Value Survey (SVS), an internally reliable self-report questionnaire commonly used to assess the theorized ten core human values (Schwartz 1992). The SVS's greatest asset is that it is now the common currency of values researchers around the world.

As impressive as the Schwartz approach to values is, it is constructed on the foundation of people's self-theories. That is, the SVS requires people to evaluate themselves along a predetermined group of 10 values that are assumed to take a specific structure constituted of specific content. Ultimately, this structure and content are imposed upon research participants by the fact that they are inherently built into the questionnaire and its scoring methods – a necessary practice for nearly all self-report questionnaires. Importantly, this is a very different approach than simply asking people for their own thoughts on the question of "What are your personal values that guide your decisions and behaviors?" Indeed, if asked this question, many people might answer "to work hard", "be faithful to my religion", or "be a good mother". Such professed values are not inherently contradictory to the SVS. Rather, the SVS lacks the ability to concretely reflect those specific values that people hold in their own personal value constellations.

An even more complex problem arises when studying the relationship between values and behaviors. Unfortunately, most studies attempting to examine value–behavior links have simply compared self-reported SVS values with other self-report attributes such as personality, likes, and dislikes. This creates a problem wherein researchers are often

ultimately exploring the relationships between different facets of people's explicit self-concepts rather than studying more organic and real-world instantiations of values and behaviors. In fact, Schwartz has pushed for researchers to explore behaviors in more detail. This undertaking seems promising and has been the focus of recent research that seeks to build a set of self-report behaviors that correspond to the values measured by the SVS (Butenko and Schwartz 2013). Unfortunately, many of the self-reported behaviors thus far have been general abstractions rather than concrete behaviors. For example, the behavioral measure for the value of "stimulation" was "change plans spontaneously", and for the value of "humility", "play down my achievements or talent."

A related issue with which all social scientists struggle is the question of how to measure behaviors efficiently and effectively. Self-reports of behaviors via forced choice questionnaires ultimately suffer from the same problem as other self-report measures: the questionnaires only contain questions that researchers think to ask. By adopting such an approach, researchers run the risk of imposing a potentially skewed, and sometimes inaccurate, structure on behavioral patterns. These are intractable features inherent to virtually all closed-format self-report questionnaires. In most cases, we would like to know what behaviors our respondents are actually doing and thinking about without relying upon questionnaire prompts. Currently, researchers are beginning to acquire greater amounts of objective behavioral measures such as buying behaviors, movement information, and even reading pattern data as the "big data" revolution continues to grow (Kolb and Kolb 2013). In the interim, however, researchers now have access to an endless stream of open-ended reports of mental life in the form of social media. A principal benefit of these reports is that they are ecologically valid and driven entirely by what people say they are doing and thinking in their own words.

The current study examines values and behaviors that emerge from open-ended text. The first of two projects relies on an online survey. This survey involved multiple randomized tasks that included 1) asking people to describe in detail the basic values that guide their lives, 2) asking people to describe the behaviors in which they engaged within the past week, and 3) participant completion of the self-reported SVS. Using a topic modeling technique called the meaning extraction method (Chung and Pennebaker 2008; Kramer and Chung 2011), values and behaviors were inductively extracted from the texts. Value- and behavior-relevant thematic factors were then compared with each other and with the SVS data.

The second project adapted the results of the first project and applied them to status updates from over 130,000 Facebook users; these data are part of the myPersonality project (Kosinski, Stillwell, and Graepel 2013). Although a relatively small number of the cases (N = 1; 260) included the SVS, the primary analyses revealed intuitive links between the MEM-derived values and MEM-derived behaviors. The work presented here, then, constitutes a proof-of concept study demonstrating the utility of relying on natural language markers of abstract psychological phenomena, including values, to better predict and understand their connections to behaviors and thought in a broader sense.

PROJECT 1: VALUES AND BEHAVIOR IN AN ONLINE SURVEY SAMPLE

To begin, we sought to determine how well the SVS captures prevalent values as described by people when discussing the things that are most important to them (i.e., their core personal values) in their own words. Additionally, we sought to explore the links between values (both from the SVS and people's free responses) and human behaviors as they manifest themselves in the real world. Theoretically, values should exhibit a discernible influence upon behaviors, including language use. As such, we expected to see

that the values reflected in a person's descriptions of their guiding principles would show relatively intuitive, predictive links to everyday behaviors. To capture this information, we designed a social survey using the Qualtrics Research Suite; the survey was then distributed using Amazon Mechanical Turk (AMT). Survey takers were presented with a series of randomized tasks that included a values essay and a behavior essay. In order to assess participants' values in their own words, they were asked to respond to the following prompt:

> *For the next 6 minutes (or more), write about your central and most important values that guide your life. Really stand back and explore your deepest thoughts and feelings about your basic values. You might think about the types of guiding principles that you use to make difficult decisions, interact with other people, and determine the things that are important in your life and the lives of those around you. Try to describe each of these values and their relationship to who you are. Once you begin writing, try to write continuously until time runs out.*

Similarly, a prompt was given with the aim of collecting natural language related to everyday behaviors. This prompt was not intended to acquire a list of all behaviors in which all participants engaged. Rather, our goal was to acquire a natural language behavioral inventory that reflected common, psychologically meaningful behaviors. The writing prompt read as follows:

> *For the next 6 minutes (or more), write about everything that you have done in the past 7 days. For example, your activities might be simple, day-to-day types of behaviors (such as eating dinner with your family, making your bed, writing an e-mail, and going to work). Your activities in the past week might also include things that you do regularly, but not necessarily every day (such as going to church,*

playing a sport, writing a paper, having a romantic evening) or even rare activities (such as skydiving, taking a trip to a new place). Try to recall each activity that you have engaged in, starting a week ago and moving to the present moment. Be specific. Once you begin writing, try to write continuously until time runs out.

Respondents were also asked to complete the Schwartz Values Survey, wherein they were asked to assign integers in the range [-1,7] to the 57 different value items of the SVS based on how important they perceived them to be as guiding principles in their own lives. With this scale, higher numbers indicate greater personal importance – responses were made using a Likert-type scale. Scores for the ten values were then calculated by taking the mean of the individual items that characterize each particular value type, with corrections being performed to address respondents' differences in use of the response scale. This step involves computing the average score for each individual across all 57 survey items, then centering each item's score around that average value (Schwartz 2009).

Tasks were presented in a randomized fashion between participants in order to minimize the potential for order effects, placing boundaries on any effects that may have been present. Participants were allowed to take as much time as needed to complete each section of the study and were encouraged to be as comprehensive as possible in their responses to the writing prompts. In order to filter out spam and careless responses, multiple "catch" items were randomly interspersed throughout the survey. These items asked users to select a particular answer that could be easily verified (e.g., "For this question, please select the third option") – participants who failed to respond to catch items were excluded from all analyses. Additionally, each of the essay writing samples was manually checked for coherence and plagiarism. Between the months of May and July,

2014, surveys successfully completed by 767 respondents (64.5% female, 77.1% Caucasian, 70.0% aged 26-54) were retained using the aforementioned criteria.

Analysis

In order to model the natural language data from participants into statistically actionable metrics, we employed the meaning extraction method (MEM). The MEM is an approach to topic modeling for natural language data that possesses demonstrated utility in understanding psychological phenomena, including both cognition (Chung & Pennebaker, 2008) and behaviors (Ramirez-Esparza et al., 2008). In essence, the MEM allows researchers to discover words that repeatedly co-occur across a corpus. When considering modest to large numbers of observations together, the cooccurrence of words can converge to identify emergent and psychologically meaningful themes.

Theme	Example Words
Faith (Positive)	God, Christian, Faith, Bible, Church
Empathy	People, treat, respect, kind, compassion
Family Growth	Family, good, child, parent, raise
Work	Work, best, hard, job, goal
Decision Making	Make, feel, decision, situation, difficult
Honesty	Honest, trust, lie, truth, loyalty
Faith (Negative)	Belief, bad, wrong, religion, problem
Social	Life, love, friend, relationship, enjoy
Growth	Life, learn, live, grow, easy
Indulgence	Money, enjoy, spend, free, change
Caring/Knowledge	Know, care, give, allow, truth
Openness	Happy, mind, open, positive, see
Knowledge Gain	Better, learn, understand, experience, realize
Principles	Guide, principle, situation, central, follow
Freedom	Strive, action, nature, personal, free
Certainty	Right, sure, strong, stand, thought

Table 3.1: Themes extracted by the MEM for the values essay writing task, Project 1.

These themes are then treated as independent dimensions of thought along which all texts can be quantified. Like most topic modeling methods, the MEM omits closed-class (function) words and low-frequency open-class (content) words to ensure reliability and validity. For the current research, we used software designed specifically to automate topic modeling and lemmatization procedures (Boyd 2014a). With the MEM approach, we identified 16 themes from the language generated during the values essay task (Table 3.1) and 27 themes from the behavior essay task (Table 3.2).

Theme	Example Words
Time	Night, Sunday, Friday, Thursday, today
Daily routine	Work, TV, shower, wake, sleep
Fiscal concerns	Need, spend, money, buy, make
Family care	Husband, school, nap, child, birthday
Chores	House, clean, laundry, cook, wash
Errands	Grocery, store, doctor, bank, dinner
Personal care	Shower, dress, brush, hair
Time awareness	Day, year, yesterday, week, hour
Gaming	Play, game, online, TV, video
Routine (meta)	Early, week, routine, activity, schedule
Media consumption	Online, listen, music, show, internet
Enjoyment	Friend, drink, weekend, party, fun
Exhaustion	Drove, slept, late, doctor, tire
Social maintenance	Friend, family, call, phone, visit
Car/bill	Car, bill, paid, afford
Information consumption	Watch, read, book, news
Yard work	Water, garden, yard, plant, mow
Relaxing afternoon	Start, enjoy, rest, afternoon, time
Car / body	Car, minute, gas, fix, gym
Task preparation	Start, coffee, begin, prepare, sit
Petcare	Water, cat, fed, feed
Secondary fiscal	Mturk, coffee, fix, mail, bank
Relaxation	Watch, movie, relax, pizza, summer
Travel	Walk, drive, park, trip, swim
Meetings	School, church, class, meeting, attend
Student	Work, job, parent, relax, hour
Momentary respite	Outside, television, cooking, bath, snack

Table 3.2: Themes extracted by the MEM for the behaviors essay writing task, Project 1.

The MEM-derived value themes capture the various semantic topics that people generate and, more broadly, tend to focus on when asked to reflect upon and discuss their values. Such themes lack the constraints of a forced choice questionnaire and, like other assessment methods, allow for nuance and variability between individuals. After performing the standard MEM procedures for theme extraction, we sought to determine how these topics correspond to the 10 values as defined in the SVS. To quantify each MEM-derived theme for individual respondents, we used word counting software (Boyd 2014b) to measure the rate of words from each theme as they appeared in each essay response. For example, an individual who used 4 "empathy" words out of 100 total words would attain a score of 4% for this theme. Following these calculations, we then correlated scores for the MEM-derived values with the values quantified by the SVS. This comparison is summarized in Figure 3.1.

	Conformity	Tradition	Security	Power	Achievement	Hedonism	Stimulation	Self-Direction	Universalism	Benevolence
Religion	●	●				○	○	○	○	●
Empathy					○				●	●
FamilyGrowth	●	●	●				○	◯	○	
Work										
DecisionMaking										
Honesty										●
NegativeReligion										
Social		●						○	○	
Growth										
Indulgence							●			
CaringKnowledge										
Openness										
KnowledgeGain	○	○	○				●	●	●	
Principles										
Freedom	○		○			●	●			
Certainty										

Figure 3.1: Relationships between SVS values and MEM-derived value themes, Project 1.

In Figure 3.1, positive relationships are denoted by black dots, negative relationships are denoted by white dots. Large dots indicate an R^2 value of >= .04, whereas small dots indicate a R^2 value of >= .01.

The established relationships among the SVS values seem to exhibit themselves here. For each of the SVS value dimensions, the correlations tend to exhibit an expected sinusoidal trend against the MEM-derived themes. Additionally, we see relatively intuitive correlations between MEM-derived values and the SVS in a way that might be expected. Peoples' use of words from the "religion" theme align well with the SVS Tradition value and fall in opposition to the SVS value of Self-Direction. We see small positive correlations between theme-score pairs such as Honesty/ Benevolence, KnowledgeGain/Universalism, and Indulgence/ Stimulation. However, we note that the correlations between the MEM-derived values and the SVS value scores are considerably weaker than would be expected were they reflecting identical constructs. Given their hypothetical measurement of the same broad construct (i.e., "values"), convergence would be expected to a rather high degree, reflected by moderately strong effect sizes; this was not the case. In other words, the ideas that people described when asked about their core personal values appear to show divergence from the top-down, theory driven set of values offered by the SVS. To illustrate the discrepancy, consider an example of one respondent's description of their core personal values. The following text is the entire description provided by a single participant, heretofore referred to as Participant Z, in response to the previously described "Values Essay" writing prompt:

> *Mainly in my life I try to maintain a moral standing with everyone I meet. I like to branch out and speak with others when they appear to be happy and in the mood*

to socialize. I try to work hard and make money in an honest fashion so that I may live a healthy and normal life. I try my best to maintain a positive attitude and outlook every day. I live life hoping for the best and looking forward instead of back.

Consider Participant Z's scores along the SVS dimensions (Table 3.3). While this person's scores along the 10 theorized value dimensions of the SVS provide no indication of any particularly strong or cohesive values, a casual reading suggests that this respondent does possess a coherent network of ideas that they believe guides their daily behaviors.

Value	Score	Value	Score
Achiev	0.03	Sec	-0.32
Benev	0.08	S-D	0.88
Conf	-0.22	Stim	-0.05
Hed	0.61	Trad	0.28
Pow	-1.72	Univ	-0.22

Table 3.3: SVS scores for Participant Z.

In this example, the SVS offers little insight into Participant Z's values, yet the quantification of their values from language appear to show some rather strong indications of their guiding principles, particularly when considered in relation to the sample's means (Table 3.4). Additionally, the MEM-derived value themes afford relatively transparent interpretation of the relative importance of each theme, even without consideration of the broader sample. These results should not be taken to suggest an inherent inferiority of the SVS. Rather, we emphasize that all self-report questionnaires designed to assess personal values would likely show similar discrepancies.

MEM-Derived Value	Respondent Score	Sample Mean
Faith (Positive)	0.00	0.04
Empathy	0.16	0.15
Family Growth	0.00	0.08
Work	0.22	0.05
Decision Making	0.05	0.06
Honesty	0.05	0.06
Faith (Negative	0.00	0.06
Social	0.16	0.15
Growth	0.27	0.12
Indulgence	0.05	0.06
Caring/Knowledge	0.00	0.05
Openness	0.11	0.05
Knowledge Gain	0.00	0.01
Principles	0.00	0.05
Freedom	-1.19	0.03
Certainty	0.05	0.02

Table 3.4: MEM-derived value scores for Participant Z.

Viewing values as constructs that inherently influence people's behavior, we also expect to see meaningful relationships between people's values and measurements of common, everyday behaviors in which they engage. To examine these links, we performed simple Pearson's correlations between the 27 behavioral themes extracted from participant behavior essays (quantified in a fashion parallel to the values themes) and values as assessed by both the SVS and MEM-derived themes (results are presented in Figure 3.2). As with the previous figure, large dots indicate an R^2 value of >= .04, whereas small dots indicate a R^2 value of >= .01. The results of this analysis show that the SVS values exhibit

low predictive coverage of themes related to everyday behaviors, yet the themes extracted from value descriptions show connections (i.e., effect sizes of R^2 >= .01) to more than twice as many common behavior topics. In other words, of the 27 behavioral themes extracted, only 6 are predicted by participant SVS scores. On the other hand, the MEM-derived value themes exhibit correlations with 14 behavioral themes.

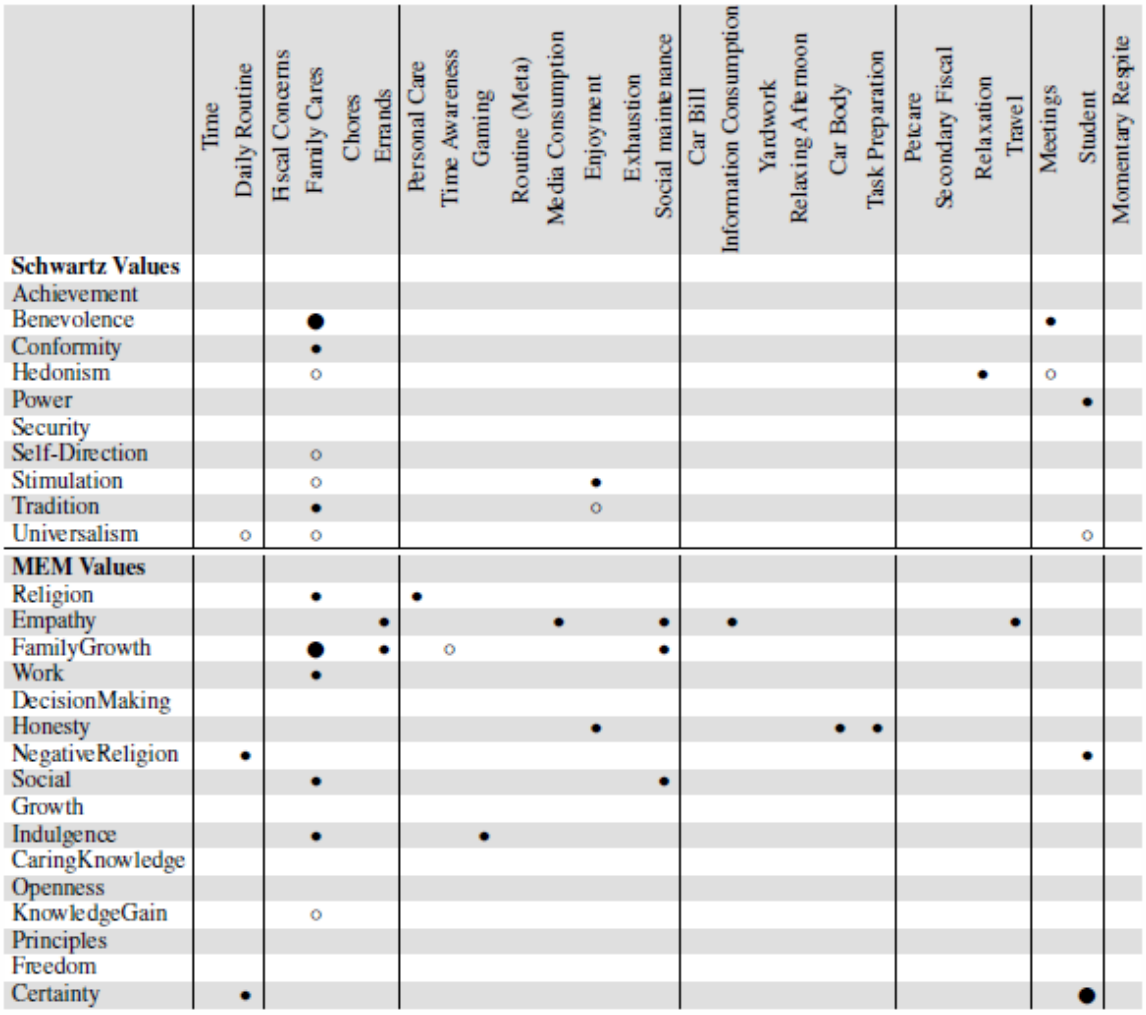

Figure 3.2: Coverage of MEM-derived behavioral themes by SVS values and MEM-derived value themes in Project 1.

The behavior themes "Relaxation" and "Meetings" were the only themes that exhibited relationships exclusively with SVS values and none of the MEM-derived value themes. Beyond these small relationships, SVS coverage of behavioral themes was in no place stronger than that afforded by the MEM-derived value themes.

In summation, the SVS dimensions are theorized to be those values that are universal and, importantly, such values are consciously accessible and able to be explicitly reported by the individual (Schwartz et al. 2012). However, in using an open-ended method for assessing a person's values where we can rely upon their own words, we see a constellation of values not captured by the top-down, theory driven approach of the SVS, which necessarily captures a limited semantic breadth. Furthermore, our language-based assessment of values exhibits better predictive coverage of an established criterion: everyday behaviors. As such, Project 1 provides further support for previous work suggesting that a person's values are predictive of behaviors. Importantly, however, we find that the network of values that are able to be captured from a person's own words appear to show predictive validity above and beyond that of a traditional self-report.

PROJECT 2: VALUES IN SOCIAL MEDIA

The primary goal of Project 2 was to conceptually replicate the results from Project 1 in a real-world social media sample. To do so, we began by examining the relationship between social media users' SVS scores and the 16 MEM-derived value topics from our original AMT sample. For this project, we used an extensive sample of social media user data is available from the myPersonality project (Kosinski, Stillwell, and Graepel 2013). This dataset consists of approximately 150,000 Facebook user's status updates. Additionally, various subsamples of these users have completed some portion of a battery

of dozens of questionnaires pertaining to personality assessment, demographics, and values.

While our AMT sample in Project 1 revealed value themes using language explicitly related to people's core values, value-laden language is also prevalent in everyday life (Chung and Pennebaker 2014). In Project 2, language pertaining to values and behaviors are not inherently differentiated, as all language was acquired exclusively from user status updates. As such, we used the MEM-derived value lexicon created within Project 1 as our "ground truth" for value-relevant words in Project 2. MEM-derived values for Facebook users were measured using word counting software (Boyd 2014b) to scan user status updates for the predetermined value-relevant words; this procedure was parallel to the language-based value quantification method described for Project 1.

To ensure reliability, all participants were required to have a minimum of 200 words used across all status updates (participants meeting criteria: N = 130,828). Those users included in the myPersonality dataset who had completed demographic surveys reported an average age of 25.3 years (SD = 11:1), and 56% identified themselves as female. Additionally, a subsample of the myPersonality dataset included Facebook users who had also completed the SVS online (N = 1,260).

Analysis

As a first step, SVS scores for Facebook users were correlated with the MEM-derived value themes as they were present in the users' status updates (Figure 3.3). Again, we see only partial coverage of value-relevant language in terms of value dimensions captured by the SVS. However, in this sample, we see a decrease in the predictive coverage of the SVS with regard to value-laden words in participant status updates. The weakened correspondence between these two measures is to be expected – unlike Project 1,

participants are not likely to be explicitly enumerating their core values. However, these results also suggest that those constructs measured by the SVS may not permeate into everyday life to the extent that researchers have typically assumed, whereas value-laden language does.

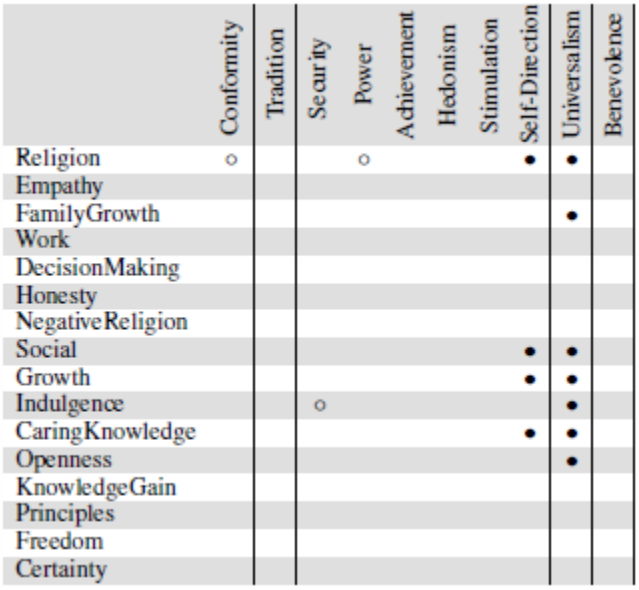

Figure 3.3: Relationships between SVS values and MEM-derived value themes, Project 2.

As with Project 1, we also sought to examine the links between Facebook users' core values and other aspects of mental life, primarily behavior. As was described for the first project, we first used the MEM to extract topical themes from the entire myPersonality corpus that met our minimum word count inclusion criteria. This procedure resulted in 30 broad themes found within Facebook user status updates (Table 3.5). A few of the behavioral themes derived from the Facebook users' language have analogs to those themes found in the AMT behavior essay responses (e.g., "Day to Day" and "Daily Routine", "Children" and "Family Care") but, in general, many of the themes derived from Facebook status updates pertain to qualitatively novel topics. Unlike the behavioral themes

from the first project, the topics in the status updates give us insight not only into what people are doing in behavioral terms (e.g., eating, studying, expressing gratitude, playing games), but also the things about which they are thinking (e.g., privacy, national issues, illness).

Theme	Example Words
Achievement	Success, courage, achieve, ability
Daily routine	Dinner, sleep, shower, nap, laundry
Going to events	Ticket, event, contact, free, tonight
Wonderful	Sky, dream, heart, soul, star
Student responsibility	Class, study, paper, homework, exam
Recreation planning	Weekend, flight, beach, summer
Religiosity	Lord, Jesus, bless, worship, pray
Eating & cooking	Soup, sandwich, pizza
Fun personality	Cute, loveable, funny, goofy
Anticipation	Amaze, excite, birthday, tomorrow
Sports	Team, game, win, baseball, football
Celebration	Birthday, Christmas, anniversary
Swearing	Ass, bitch, dick, fucker
Internet movies	Watch, movie, youtube, episode
Privacy declaration	Settings, information, account, privacy
Nationalism	Liberty, America, nation, flag, unite
Parental protection	Childhood, violence, campaign, abuse
Cancer support	Cancer, patient, cure, illness
Musicianship	Band, guitar, rehearsal, perform
Friendship gratitude	Cherish, friendship, post
Farmville	Farmville, stable, barn, gift
Group success	Succeed, hug, cheer
Web links	http, org, php
Concern for underprivileged	Elderly, homeless, veteran
Proselytizing	Deny, believer, Christ, heaven
Celebrity concerns	Marriage, Britney, Spears, Jesse
Severe weather	Severe, thunderstorm, tornado, warning

Table 3.5: Themes extracted using the MEM on Facebook status updates.

Importantly, many of the behavioral themes that were extracted from the corpus included words that were also found within the MEM-derived value themes found in

Project 1. Many behaviors in which people engage will necessarily be value-laden to some degree, however, we sought to minimize effect size inflation due to shared word use between Project 1's MEM-derived value themes and Project 2's MEM-derived behavioral themes. As such, words that appeared in both sets of themes were systematically omitted from the behavioral themes prior to quantification. As with value-relevant words, each Facebook user's entire set of posts was then quantified along each MEM-derived behavioral dimension using the same word counting approach described above.

Finally, we performed an analysis parallel to that described for Project 1 in order to explore the degree to which the language-derived value themes and SVS value scores corresponded to the self-described behaviors and ideas present in Facebook users' status updates. We emphasize two primary aspects of the results, presented in Figure 3.4. First, we again see a conceptual replication of Project 1 in terms of value-behavior relationships. Scores from the SVS appear to show little correspondence with the actual behaviors and ideas that our sample of Facebook users share with others, whereas language-derived values show considerable and consistent relationships with behavioral topics. Second, whereas the SVS appears to correspond to rather narrow bands of behavioral themes, the language-derived values show extensive coverage of behaviors in predictive terms. In other words, the results from Project 2 not only conceptually replicate the results from Project 1, but demonstrate the applicability of the language-derived value themes to a completely new set of themes pertaining to the common thoughts and behaviors of social media users in the real world.

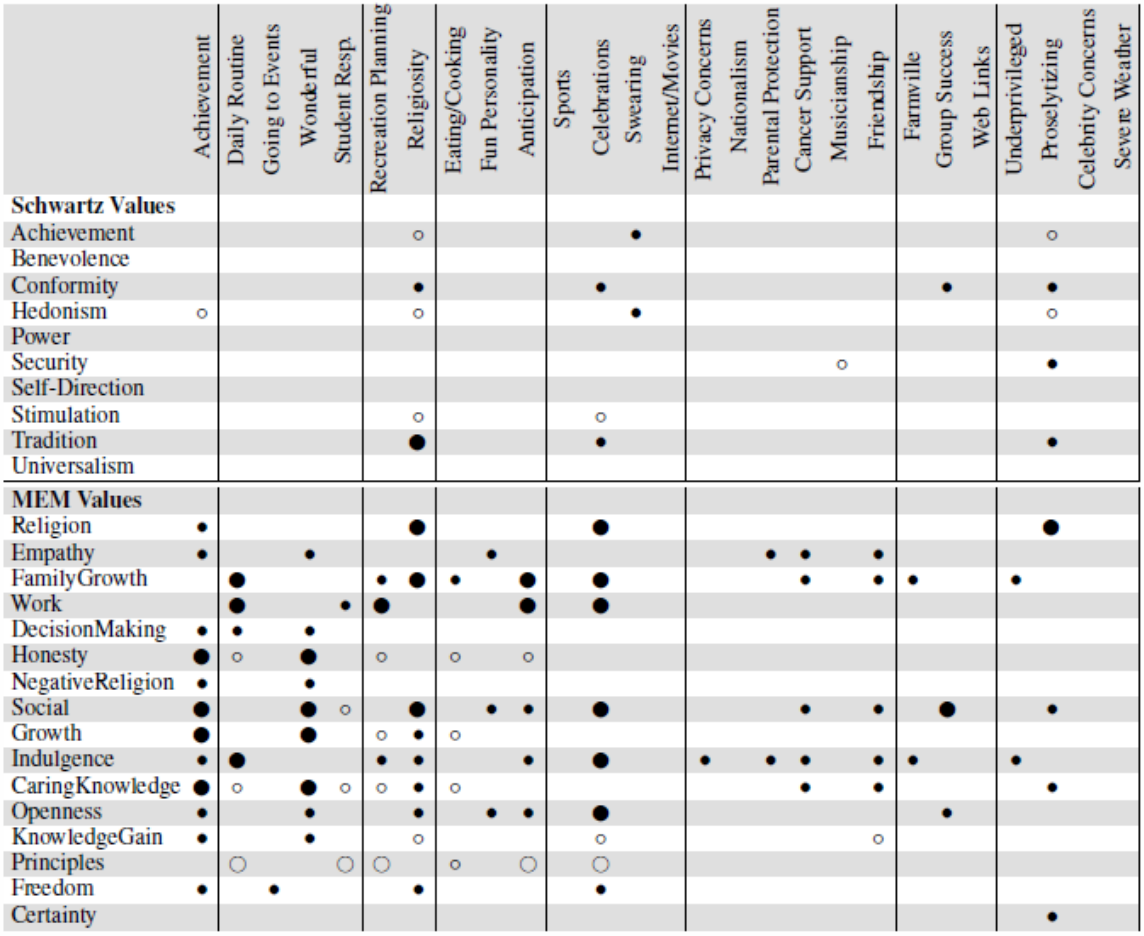

Figure 3.4: Coverage of behavior MEM themes by SVS values and value MEM themes, Project 2.

CONCLUSIONS

We have collected and analyzed one new, crowd-sourced dataset and one archival social media user dataset in order to better understand the relationships between people's values and their behaviors using a natural language processing approach. We found that the widely-adopted set of values that are measured by the SVS provide substantially less predictive coverage of real-world behaviors than a set of values extracted from people's own descriptions. Simply asking people what is important to them turns out to be a more informative method for answering the question of what values are, and the simple word

counting approach appears to be a viable method for value quantification. Using this approach, we examined a large-scale social media data set to explore whether the language of values would continue to exhibit relationships with the ideas and behaviors that people share in their Facebook status updates. Results offer consistently strong support for language-based value–behavior links.

It is our hope that this study will open more doors to future work in values research. A new set of values has been identified, along with a method that allows for the simple, intuitive lexical representation of values. These methods can be used to study the values of various groups of people across various platforms, languages, time, and space. We note that this approach requires that a large enough body of text be collected for successful research. However, this is easily achieved by using more social media data, blog data, and other forms of prevalent data available in the current big data atmosphere. This approach may also facilitate further exploration of the relationships that exist between values and behavior by encouraging more fine-grained computational models.

Beyond Values

We have shown here a single case in which natural language data provided a more clear picture of people's cognitive and behavioral processes than data collected from a traditional and widely used self-report survey. Additionally, we have demonstrated that the information extracted from natural language exhibited more links (both in terms of quantity and diversity) with behaviors and thoughts than a standardized self-report measure. However, we advocate that the general approach that we have used for the current studies can also be applied much more generally. Indeed, many of the social and psychological phenomena studied using social media are conceptually abstract and difficult to distill into valid metrics. While the standard approach to studying such phenomena is to rely on

gathering self-report data in the form of forced-choice questionnaires, this process often requires the collection of data beyond what is already available via social media and may often serve as insufficient "ground truth" when attempting to capture psychology as it exists in the real world.

As described in the current work, we emphasize that already-existing, organically generated social media data can exhibit greater predictive strength for human behaviors and a more dynamic structure than that imposed by closed, forced-choice questionnaires. Additionally, data at the "big data" level are often only available in the form of natural language. In such cases, we have demonstrated that psychological "ground truth" can still be attained, allowing researchers to explore human psychology under conditions where diverse forms of data are unavailable. Finally, the methods described here allow for the inference of many different psychological phenomena from the same data, including the core three components of human psychology (i.e., affect, cognition, and behavior). It is our aim to demonstrate with the work presented here that language is an incredibly flexible form of data that can be used to many great purposes.

Chapter 4: Mental Profile Mapping

INTRODUCTION

Authorship attribution is, broadly speaking, the process by which works of unknown or disputed origins are investigated to determine their history. In the past, various approaches have been taken to establish authorship information about questioned documents, including things like chemical analysis of physical documents, identifying idiosyncratic spellings or phrases (i.e., "stylometry"), and even the formation of subjective, holistic impressions of the contents of a text for fit with a specific authorial candidate (e.g., "this just *feels* like Shakespeare"). Regardless of the specific methodologies, all authorship attribution tasks are inherently forensic in nature: by establishing patterns common to a known author or authors, it is hoped that the general history and origin of texts with unknown authorship can be partially, if not fully, reconstructed. The past 2 decades have seen an explosion of new methods in the world of authorship attribution, particularly those that employ statistical modeling of language to determine authorship likelihoods (see Juola, 2008).

Despite the recent boom in sophisticated text analytic authorship attribution methods, however, tensions often exist in forensic settings where impenetrable algorithms are given free reign over authorship questions at the exclusion of intuitive, digestible, and human insights (e.g., see Solan, 2013). Many people tend to have an "algorithm aversion", or a distrust of opaque algorithms that cannot be easily interpreted by laypersons (Dietvorst, Simmons, & Massey, 2014; Promberger & Baron, 2006). In simple terms, most people often find it difficult to place blind trust in an opaque, cold probability score generated by an algorithm, especially when the processes by which results are generated are poorly understood.

Skepticism may be particularly pronounced when it is important for individuals to be able to develop an intuitive understanding of forensic methods and their results (e.g., Bromby, 2011). Complex machine-learning methods may not only jeopardize a layperson's ability to interpret the results of forensic text analyses, but also the ability of expert researchers themselves to adequately understand the process by which results are attained (e.g., by interpreting an algorithm's resultant model).

In many settings, then, it may be necessary to strike a compromise between sophisticated analytic techniques and deeper, actionable insights that lend themselves to meaningful interpretation. In the case of authorship attribution tasks, this can take the form of methods that create information extending beyond probability statements, such as verifiable idiographic data about an author. For example, rather than a result reading something like "Thomas is 85% likely to be the author of this document", a more balanced analytic approach may also provide extra information in the form of "The author of this document also appears to be impulsive, authoritative, and extraverted, which matches observer reports of Thomas's personality".

Notably, methods in the realm of psychological text analysis have advanced by leaps and bounds separate from, yet in parallel with, the proliferation of authorship attribution methods. Much recent work in the psychological sciences has found that the psychological properties of an author can be accurately captured using automated text analysis procedures. Research spanning hundreds of labs around the globe have repeatedly found that various categories of language are direct reflections of personality-relevant psychological processes (see Tausczik & Pennebaker, 2010; Boyd & Pennebaker, 2015b), suggesting that a person's mental life can be adequately modeled from modest language samples. In other words, various dimensions of a person's mental world, such as their emotional, social, and cognitive tendencies, can be captured indirectly, yet accurately, by

measuring psychologically relevant patterns in language. Moreover, given the trait-like properties of psychological measures of language (e.g., Pennebaker & King, 1999), it has been found that individuals are uniquely identifiable by the very psychological traces in their language (e.g., Boyd & Pennebaker, 2015a).

Such discoveries are paving the path for new combinations of computer science and psychology in a forensic space, however, computational approaches to psychological forensics are currently in their infancy. Many methodological gaps still exist for common tasks such as authorship attribution within each field separately, and virtually no methods exist that successfully combine the two fields to resolve these problem areas. Simply put, most authorship attribution methods are either wholly computational or wholly psychological in nature; these two fields seldom cross paths, yet have great potential for mutual benefit.

The current study brings together computational forensics with psychological forensics by introducing a new method for single-candidate authorship attribution, named *Mental Profile Mapping*, which aims to fill several critical methodological gaps. By combining these two disparate fields into a unified approach, critical gaps are filled within each field, as well as in the broader authorship attribution literature.

Contemporary Authorship Attribution Methods: Background and Gaps

The majority of modern authorship attribution methods use statistical analyses that fall under the umbrella of *supervised machine learning* (SML). SML methods allow a computer system to be "trained" on data where concrete outcomes are known. In practice, trained models can then be used to predict outcomes in new, previously unseen data. For example, if a SML algorithm is trained on a collection of images with known faces, it can

be used to accurately identify familiar faces in a novel image as a function of what it has previously learned (e.g., Bhele & Mankar, 2012).

The power of SML methods is in their ability to discriminate between multiple known outcomes that exist in a training dataset with extraordinary accuracy. The appeal of SML in authorship attribution tasks is readily apparent: if a system can be trained on the language patterns of various known authors, a work of questionable authorship can be statistically assigned to one of those authors. These types of problems are known as "multiple-candidate" or "closed-class" problems in authorship attribution – that is, a work is known (or strongly believed) to originate from one of *N* specific authorial candidates.

A considerably more difficult problem in authorship attribution is one of "single-candidate" attribution tasks. In single-candidate problems, the authorship question boils down to "did *Person X* write this text?". Currently, several methods have been proposed to address single-candidate problems, however, these methods typically convert the single-candidate question to one of multiple candidates, for example, by introducing "impostors" who are known *a priori* to have not written the work in question (e.g., Koppel & Winter, 2014). These methods are useful when comparable data is readily available/accessible, however, they are less practical in cases where data is limited or of a unique variety. From a psychological perspective, such methods are also lacking deeper insights. Like other authorship attribution methods, most single-candidate attribution methods typically rely on word distributions that provide no further information into authorship beyond results in the form of probability outputs.

To illustrate this last point, consider a hypothetical scenario wherein an unknown author has left behind an unsigned admission of guilt for arson. The police suspect Joseph in the case and have obtained several other of Joseph's "baseline" writings. The question, then, is whether it can be determined that the person who wrote the baseline writings (i.e.,

Joseph) also authored the admission of guilt. A standard authorship attribution analysis would decay all texts into a series of high-dimensional vectors based on words and word properties (e.g., part of speech, repetition, etc.), then statistically determine the likelihood that they came from the same source using classification algorithms. Particularly when conducting complex tasks such as authorship attribution, advanced statistical and machine learning methods typically preclude any psychological understanding of the person who actually created a text (see Boyd, in press), which may be of great importance in forensic and courtroom settings.

Now imagine the same scenario described above, yet an analyst employs a *psychological* approach to authorship attribution rather than a purely statistical approach. Using psychological text analysis procedures, an analyst would be able to extract information about Joseph from his baseline texts (e.g., "Joseph's baseline texts are indicative of someone who is generally extroverted, honest, and unanalytical."). Furthermore, Joseph's friends, neighbors, and family also agree that these traits are an apt description of his general personality. In this scenario, the statistical information of standard authorship attribution analyses can be combined with the additional psychological information (e.g., the observer reports of Joseph, the psychological patterns embedded Joseph's texts) to form a more robust analysis and interpretation of the findings. If the psychological patterns extracted from the admission of guilt are indicative of someone with the same general psychological traits as Joseph, then the forensic account of the admission is strengthened.

The above example is a conceptual demonstration of why it is useful, then, to strike a compromise between computational and psychological methods of authorship attribution that satisfies 2 criteria: 1) valid authorship attribution frameworks must be underpinned by empirical, reproducible, and quantifiable *methods* rather than intuition or subjective

judgments (computational/statistical perspective), and 2) ideally, an authorship attribution analysis should provide *results* that can be interpreted in the context of relevant psychological information such observer reports of an individual's traits, behaviors, and mental profile (psychological perspective).

Such a compromise allows the rigor of advanced authorship attribution methods to be paired with quantification and analytic methods that lend themselves to a deeper, meaningful interpretation from a psychological perspective.

Modern Psychological Authorship Attribution

Recent research has found that individuals can be differentiated based on their unique psychological composition, which can be reliably measured through language use. Boyd and Pennebaker (2015a) found strong support for this idea by demonstrating that authors could be robustly and reliably differentiated with near-100% accuracy using exclusively psychological measures of language. Unlike most authorship attribution techniques, which are often based on more atomic linguistic measures (e.g., distributions of short phrases, spelling errors, etc.), Boyd and Pennebaker's (2015a) methods lent themselves to a scientific, psychological analysis of an author by identifying their unique psychological attributes.

The ability to engage in follow-up interpretations of language patterns stands in stark contrast to most authorship attribution methods. Furthermore, their results were also able to be compared with observer reports of authors' personalities, dispositions, and even behavioral events. In their work, Boyd and Pennebaker's language-based analysis of authors' psychological profiles showed strong convergence with other psychological data, strengthening the forensic account provided by their authorship attribution analysis. Lewis Theobald, for example, exhibited language patterns consistent with a highly analytic yet

socially cold and confrontational personality. Indeed, observer reports of Theobald mirrored the insights gained from language analyses, as he was known to be brilliant but quarrelsome with colleagues and had few to no close friends.

Nevertheless, the methods used in Boyd and Pennebaker's (2015a) study of psychological authorship attribution possess some drawbacks common to non-psychological methods. For example, the methods used in their work are only useful for multiple-candidate attribution tasks – single-candidate problems are not typically able to be solved by means of classification tasks alone. Therefore, while Boyd and Pennebaker's analysis represents a promising first step towards the unification of computational and psychological sciences in the domain of forensic analyses, more work is required to extend and expand this new type of approach into uncharted territory.

Current Study

In the current study, a new method of authorship attribution and, more broadly, psychological profiling is introduced. This new method, named *Mental Profile Mapping*, aims to address multiple critical gaps in the forensic space of authorship attribution. Namely, there currently exists a marked lack of empirical, theory-based psychological methods for single-candidate authorship questions. Currently, no explicitly quantitative methods for single-candidate authorship attribution exists in the field of psychology. The methods underlying Mental Profile Mapping not only rely exclusively on psychological features that can be measured from language, but the results of this method can be interpreted in a psychological manner through subsequent decomposition and analysis.

The current study first demonstrates the results of a high-power and validated authorship attribution method hailing from computer science, known as "unmasking", which is applied to a test case – the works of Aphra Behn. After baseline results are

established in the Behn authorship problem space, Mental Profile Mapping methods are introduced. Results from the new analytic approach are then compared and explored in the context of the unmasking results and other psychological information.

METHODS

Methodological Test Case: Authorship Attribution with the Works of Aphra Behn

For the current study, the plays of Aphra Behn were used as the focal data source for analysis. Aphra Behn is regarded as one of England's first female playwrights and among the 17th century's most influential dramatists. Due to the controversial nature of Behn's writing, as well as gender politics of her own and subsequent periods, much of her work and legacy was suppressed for a considerable stretch of literary history between her death and the latter half of the 20th century (Spencer, 2000). However, Behn's legacy has rapidly become the object of much study and interest in modern literary research (see Todd, 1998).

Born in 1640, shortly after William Shakespeare's death, Behn's childhood is shrouded in mystery, and much of her early life history appears to be obscured either intentionally or due to extraneous factors (Hughes, 2001). Behn served as a spy during the Second Anglo-Dutch War prior to becoming a skilled and well-performed playwright, and served several tours during wartime (O'Donnell, 2004). Much of Behn's work was highly successful during its time, and Behn's death in 1689 caused an upsurge in the demand for published copies of her work.

Profiteers and publishers of the time met the demand for copies of Behn's work by compiling her plays and commissioning several rounds of printing. In particular, Charles Gildon and Gerald Langbaine helped to fuel additional demand through their (perhaps sensationalized) biographical accounts, occasionally mixed with praise of her personality

and exploits. In recent years, doubts have been cast regarding the true origin of some of Behn's posthumously-published plays, particularly those in which Gildon was involved. Gildon was an occasional literary forger and writer, and his involvement in the publication and dissemination of Behn's works has raised suspicions about whether the posthumous publications are authentic, or merely opportunistic forgeries (see Spencer, 2000).

The works of Aphra Behn are a suitable test case for authorship attribution methodologies for several reasons. First, Behn's plays are composed of tens of thousands of words, which makes them suitable for extracting stable language patterns within each work. Additionally, lengthy writings such as Behn's plays allow for highly reliable psychological measurements to be performed via automated text analysis (see Pennebaker, Boyd, Jordan, & Blackburn, 2015; Boyd, in press).

Furthermore, it is helpful that plays by several other authors surrounding Behn's era, both before and after her life, are readily available in machine-readable format and exist in the public domain, providing accessible comparison cases for testing. Lastly, scholars have painstakingly assembled thorough timelines of events, records, and other chronological information for the life and works of Behn (see O'Donnell, 2004). Importantly, this last factor allows the results of a psychological authorship attribution analysis to be tentatively compared to details of Behn's life, facilitating a convergence of empirical methodology with external psychological information.

Setting an Authorship Expectation Baseline: The "Unmasking" Analysis

Before introducing the Mental Profile Mapping approach, it is important to first establish basic results within the Aphra Behn authorship problem space using an established, high-powered, and validated technique. By setting baseline expectations, the Mental Profile Mapping analysis can be more thoroughly evaluated for convergent validity.

Essentially, if results from the new method are comparable to that of an established computational method, it becomes easier to place faith in the results of both analytic tactics. For the current analysis, a powerful modern authorship attribution technique from the computational sciences was selected to set a baseline for expectations pertaining to authorship results – this method is known as "unmasking" (see Koppel, Schler, & Bonchek-Dokow, 2007).

A brief description of unmasking

The unmasking method, like many other authorship attribution methods, requires several texts by multiple known authors to determine the origins of a questioned work. However, unlike similar methods that convert single-candidate problems into multiple-candidate problems, the unmasking method occupies a unique hybrid space between the two types of problems. In essence, unmasking operates by using machine learning methods to model the "depth of difference" between an author's known works and those texts by other authors. Subsequent unmasking stages then use this information to classify unknown texts for authorship. The results from the unmasking method come in a fairly straight-forward form: a questioned work either is or is not a match with a given author (with a given probability). Because of the specific process by which unmasking operates, it is a uniquely "open-class" approach to authorship attribution that is suitable for the current comparison.

The unmasking method involves several stages that coalesce into a meta-learning algorithm designed for authorship attribution. The logic of the unmasking method is both quite clever and conceptually simple, yet rather complex in its execution. The underlying idea behind the method is this: if we select one work at random out of an author's complete works, it would be rather easy to use machine learning methods to differentiate the selected

work from the rest of the author's works. However, if we were to iteratively remove a handful of those features that best differentiate the selected work from the others, the differentiation process becomes increasingly difficult with each iteration.

For example, *The Tommyknockers* by Stephen King may have "superficial" differences from his other works, such as different characters and themes, but there will also be several linguistic patterns that are constant throughout his works by virtue of the fact that he himself is created the prose. As one gradually strips away these superficial features (e.g., themes, settings, characters), all of his works become increasingly difficult to differentiate – this iterative drop in accuracy creates a "prediction degradation curve". In contrast, if the same process is performed to compare *The Tommyknockers* with the works of several other authors, King's novel will still be able to be uniquely identified rather easily. Even after the removal of superficial differences between *The Tommyknockers* and works by other authors, King's imprint on the book's language remains distinct from other people's writings. In other words, the "fingerprint" of King's language is ultimately still unique enough to differentiate his work from novels written by other people – there is very little prediction degradation, resulting in a rather shallow (or even flat) curve.

The essence of the unmasking method lies in the generation of these prediction degradation curves. The degradation curves, with some additional information, are used in a meta-learning algorithm to identify when works do (or do not) belong to any given author. In practice, unmasking uses support vector machine (SVM) models with linear kernels. Texts by each author in a corpus are initially tested to see how well they fit into an author's own corpus versus an amalgam of "different author" works. Initially, results are typically very strong – works by a given author share enough linguistic features to stand apart from the works of all authors.

The basic unmasking process is repeated several times (e.g., $f = 10$). However, during each iteration, a select number of features (e.g., $k = 2$) that best differentiate each work from either the same author or those of others are removed, thus negatively impacting SVM classification accuracy. The reasoning behind this approach is that prediction degradation will occur much more quickly for same-author works than different-author works. The results of each iteration (i.e., "fold", or *f*) are stored and later used in a separate SVM model (i.e., the meta-learning model) that can identify whether questioned works belong to a given author as a function of their prediction degradation curves.

Data

Aphra Behn corpus

A collection of works by Aphra Behn was provided by Melanie Evans, Elaine Hobby, and Claire Bowditch (e.g., Bowditch & Hobby, 2015; Evans, 2016) as part of an upcoming compilation of Behn's works. All texts were provided with modernized spellings and had extraneous text (e.g., dramatis personae, title pages, prefaces) removed. The list of plays included in the current study is presented in Table 4.1.

The Forc'd Marriage	The Rover, Part I*	The Roundheads
The Amorous Prince	Sir Patient Fancy	The False Count
The Dutch Lover	The Feigned Courtesans	The Young King
Abdelazer*	The Rover, Part II*	The Emperor of the Moon
The Town Fop	The City Heiress	The Lucky Chance
The Widow Ranter		

Table 4.1: Aphra Behn plays included in the current analyses.

Note: Adaptions of other people's work are denoted with an asterisk (*).

Works of questioned authorship

In addition to the verified works of Aphra Behn, 5 plays of questioned authenticity were provided: *The Debauchee, The Woman Turned Bully, The Counterfeit Bridegroom, The Revenge,* and *The Younger Brother.* All works of questioned authorship were prepared/cleaned in a manner analogous to that described above.

These 5 plays were selected as works of potential Behn authorship along several criteria. First, most of these plays were published between 1675 and 1680, a period during which Behn was at her most prolific. While some of the questioned plays do not bear Behn's name as an author, most of them are thematically congruent with her known authored works. Two of these plays (*The Debauchee, The Revenge*) were marked as possible Behn works due to their use of prostitutes as central characters, whereas *The Woman Turned Bully* and *The Counterfeit Bridesgroom* feature women cross-dressing as men to achieve financial independence – both tropes being markedly common in the works of Behn.

Finally, the last questioned play (*The Younger Brother*) was posthumously published in 1696. While this play was, in fact, attributed to Behn at the time of its publication, it was prefaced by Charles Gildon, who claimed to have made his own edits to Behn's original work and is commonly viewed as an unreliable source. Given that Behn's works were a profitable commodity at the time of its publication, this final play can be seen as suspect for true Behn authorship. For additional discussions on the matters of authorship details surrounding these works, please refer to Spencer (2001) and O'Donnell (2004).

Supplemental unmasking corpus: Preparation and analysis

Works by several additional playwrights were collected into a separate corpus as a part of the the "unmasking" analysis procedures. Supplemental playwright data was

included in the "unmasking" corpus that included the works of William Shakespeare (35 plays), Christopher Marlowe (5 plays), John Fletcher (9 plays), Lewis Theobald (12 plays), William Rowley (1 play), and Thomas Dekker (13 plays). Works by these additional authors were selected based on several criteria, primarily their ready availability, as well as their genre similarity to the plays of Behn, and the fact that all supplemental playwrights lived in England within a century of Behn (years of birth and death range from 1564 to 1744; Behn lived from ~1640 to 1689).[3] All supplemental works were gathered independently but were prepared in a manner similar to the works of Behn by means of spelling modernization, extraneous information removal, and so on (see Boyd, 2014).

Text Analysis Method

All plays by all authors were analyzed using LIWC2015 (Pennebaker, Booth, Boyd, & Francis, 2015). The LIWC2015 software codes texts for words belonging to ~80 psychological dimensions of language previously found to be related to emotions, social and cognitive processes, and attentional processes. The LIWC2015 application operates by calculating the frequency of words belonging to each category, then dividing by the total number of words. For example, if a text has 1 positive emotion word (e.g., "pleasure") out of 10 total words, the text is scored as 10% for positive emotion words. This method has been extensively validated in research the past two decades, and the LIWC categories are often used in authorship attribution studies (see Koppel, Schler, Argamon, 2008; Pennebaker, 2011). LIWC features have also been found to be as useful in psychological

[3] It is important to note that, unlike the majority of other authorship attribution methods, the specific authors selected for this unmasking analysis are not of particular importance. Ultimately, the unmasking approach does not cleanly distill into a traditional multiple-candidate authorship problem that *must* select one of the candidates as the "true" author (i.e., "Text X *must* have been written by one of these authors – therefore, which author most likely wrote Text X?"). Instead, supplemental authors are simply used as a sounding board for the purpose of modeling same- versus different-author prediction degradation curves.

analyses of authorship attribution as standard n-gram distributions (Boyd & Pennebaker, 2015a).

Plays by all authors were segmented into chunks of ~250 words and analyzed using LIWC2015 (Pennebaker, Booth, Boyd, & Francis, 2015) in accordance with unmasking procedures. The unmasking procedure was then applied to the resulting 8,066 play segments in the precise manner outlined in Koppel, Schler, and Bonchek-Dokow (2007). This method was performed using 10 folds in conjunction with a *k* parameter set to 2 (i.e., 4 features dropped per iteration). Additional prediction degradation curve data was included in the overall feature set, including linear and quadratic polynomial betas and $i{+}f_{\{1,2\}}$ degradation Δ scores.

RESULTS AND DISCUSSION

The unmasking method performed extremely well with the psychological data generated by LIWC for each author (see Table 4.2). The meta-learning algorithm showed strong performance for correctly identifying same-author versus different-author degradation curves across the entire dataset (accuracy = 93.96%; kappa = .74). Results were equally strong when considering same-author versus different-author curves for Behn alone (class-balanced accuracy = 88.89%; kappa = .78).

Ultimately, the goal of the unmasking method is to determine whether works of unknown origins were created by a known author. Typically, the unmasking procedures are useful for considering a single author candidate (e.g., "Did Aphra Behn write *X*?"). However, the results of this method can be looked at more broadly as well, ensuring that other authors who are known to not be "true" candidates are also ruled out. Full results from the unmasking analysis are presented in Table 2. Plays highlighted in green indicate a play's match with a playwright's psychological signature. Plays highlighted in light red

Questioned Play Title	Comparison Author	Authorship Match	Result *p*
The Counterfeit Bridegroom	Behn	Different	82.08%
The Counterfeit Bridegroom	Marlowe	Different	99.83%
The Counterfeit Bridegroom	Fletcher	Different	99.40%
The Counterfeit Bridegroom	Rowley	Different	98.99%
The Counterfeit Bridegroom	Theobald	Different	99.97%
The Counterfeit Bridegroom	Dekker	Different	98.61%
The Counterfeit Bridegroom	Shakespeare	Different	96.05%
The Debauchee	Behn	Different	94.41%
The Debauchee	Marlowe	Different	99.98%
The Debauchee	Fletcher	Different	95.07%
The Debauchee	Rowley	Different	99.71%
The Debauchee	Theobald	Different	100.00%
The Debauchee	Dekker	Different	99.52%
The Debauchee	Shakespeare	Different	98.32%
The Revenge	Behn	Match	99.98%
The Revenge	Marlowe	Different	99.72%
The Revenge	Fletcher	Different	98.10%
The Revenge	Rowley	Different	99.48%
The Revenge	Theobald	Different	99.89%
The Revenge	Dekker	Different	94.90%
The Revenge	Shakespeare	Different	96.88%
The Woman Turned Bully	Behn	Different	88.53%
The Woman Turned Bully	Marlowe	Different	99.93%
The Woman Turned Bully	Fletcher	Different	97.80%
The Woman Turned Bully	Rowley	Different	99.87%
The Woman Turned Bully	Theobald	Different	99.94%
The Woman Turned Bully	Dekker	Different	99.94%
The Woman Turned Bully	Shakespeare	Different	87.51%
The Younger Brother	Behn	Match	55.10%
The Younger Brother	Marlowe	Different	99.93%
The Younger Brother	Fletcher	Different	99.91%
The Younger Brother	Rowley	Different	99.58%
The Younger Brother	Theobald	Different	99.69%
The Younger Brother	Dekker	Different	99.55%
The Younger Brother	Shakespeare	Different	99.11%

Table 4.2. Results from the unmasking analysis.

indicate a play's non-match with supplemental playwrights. Plays highlighted in bright red indicate a play's non-match with Behn's psychological signature. Unmasking data was created by averaging results across 20 randomized iterations.

The results of the unmasking procedure identified only 2 of the questioned plays as having been authored by Behn: *The Revenge* and *The Younger Brother*. All remaining questioned plays were identified as extremely poor authorial fits for Behn, as well as the supplemental authors. It is also quite promising that none of the 5 questioned plays were identified as having been written by any of the "supplemental" playwrights (Shakespeare, Fletcher, Theobald, Dekker, Marlowe, and Rowley).

Having established a baseline results and expectations for the questioned plays in the current authorship space, a more thorough and thoughtful consideration of the Mental Profile Mapping procedures can be conducted and described below. Whereas the results of the unmasking method are fairly stark and closed to deeper interpretation, the methods described in the next section facilitate a psychological analysis of authorship attribution results.

MENTAL PROFILE MAPPING: A NEW AUTHORSHIP ATTRIBUTION METHOD

Underlying Concept

As a way to tackle the single-candidate authorship problem from a psychological perspective, this study introduces the concept of Mental Profile Mapping (hereafter denoted as "MPM"). The underlying concept of MPM is fairly simple: by assuming that several repeated psychological measurements are indicative of the same individual, as does any measure of personality (Funder, 2015), methods can be developed that facilitate the detection of outliers, or observations that fit poorly with the others. Because people show natural variation across time, such a method will want to look for outliers on not just

isolated, singular psychological process (e.g., social orientation, emotional state) but instead for outliers across a whole battery of psychological measures. In other words, the MPM is primarily designed to help identify violations in the broad consistency of a person's traits, including language-based measures of psychological processes. If an observation is different along not just one psychological measure, but instead more generally, suspicion as to the origins of that observation can be reasonably raised.

As an analogy, imagine that you receive an occasional e-mail from a coworker, Margaret, perhaps once or twice a month. Over time, you begin to develop a sense Margaret's personality in general terms. In her e-mails, Margaret seems to be a very positive person, keenly tuned in to the "here and now", and often mentions her personal life (e.g., her leisure activities). While Margaret does not always say the same thing – she sometimes talks about the weather instead of her bowling league, and seems less positive on some days relative to others – there is an overarching consistency to her personality that emerges in her communications.

One day, your supervisor tells you that she has received an e-mail from an unknown address, but she thinks that she knows who sent it. Your supervisor forwards you the text of the e-mail and asks "Does this e-mail look like it came from Margaret?". As you read the e-mail, you might check to see if all of the criteria for Margaret's psychological profile are met. You notice that the e-mail is generally positive and very present-focused. Additionally, there is a comment about recent leisure behaviors – the e-mail's author states that they went to a concert this past week. Given that this e-mail appears to fall into the general constellation of Margaret's mental universe, you may tentatively conclude that this e-mail does appear to fit Margaret's bill for authorship.

The principal concept underlying MPM is fairly simple and parallels the example given above. In order to establish a typical psychological pattern for an individual, several

psychological measures must be gathered from different time points. Just like the various qualities of Margaret's writing (e.g., tone, time-orientation, content), each psychological measure – in this case, LIWC-based measures – will show some variation over time yet still be reflective of the underlying source. Just as you noticed consistencies in Margaret's style, similar consistencies will emerge for an individual in their psychological processes (e.g., emotions, social processes, attentional processes). Under the assumption that all measurements were generated by the same person, outliers become suspect and merit further investigation.

Mental Profile Mapping: Quantification and Statistical Methods

By pairing psychological text analysis with a multivariate distance measure, Mahalanobis distance (Mahalanobis, 1936; Schinka, Velicer, & Weiner, 2003), it is possible to statistically conduct the type of analysis described above. Ultimately, the goal is to assess the distance of several types of psychological processes from their respective centers for a given individual, then collapse all distance scores into a single metric that can be used to better understand the overall patterns. For example, if a person's emotional state is measured 20 times, it is possible to establish the average, or "center", of their emotional states. The same could then be done with the same person's social orientation, decision-making tendencies, and so on, resulting in psychological centers for each of the different types of psychological measurements.

Unlike other distance measures, such as Euclidean distance, the Mahalanobis distance statistic is explicitly designed to handle inter-correlations among multivariate data. From a psychological perspective, the fact that Mahalanobis distance accounts for the covariance structure of the data allows a more meaningful distance metric to be calculated relative to other distance metrics. In other words, the mathematical underpinnings of the

Mahalanobis distance statistic is well-suited to handle the fact that psychological subprocesses are non-independent and may demonstrate overlap. For example, positive and negative emotions are not perfectly orthogonal (Posner, Russell, & Peterson, 2005), nor are interpersonal motives and behaviors (Horowitz et al., 2006) – interdependencies such as these exist across several of the LIWC psychological measures (Pennebaker, Boyd, Jordan, & Blackburn, 2015).

Psychological Process	**LIWC2015 Variables included in Distance Calculation**
1: Style	Analytic, Clout, Authentic, Tone
2: Complexity	Analytic, Sixltr
3: Function Words	i, we, you, shehe, they, ipron, article, prep, auxverb, adverb, conj, negate, interrog
4: Emotional	affect, posemo, negemo, anx, anger, sad
5: Social	social, family, friend, female, male
6: Cognitive	cogproc, insight, cause, discrep, tentat, certain, differ
7: Perceptual	percept, see, hear, feel
8: Biological	bio, body, health, sexual, ingest
9: Motivational	drives, affiliation, achieve, power, reward, risk
10: Temporal	focuspast, focuspresent, focusfuture
11: Relational	relative, motion, space, time
12: Personal	work, leisure, home, money, relig, death
13: Utterances	informal, swear, assent, nonflu

Table 4.3: Psychological process compositions, where each process consists of several subdimensions that are factored together when calculating distance metrics.

For the current MPM analysis, LIWC2015 was used to extract psychological data from all source texts mentioned in the unmasking section above. All LIWC measures were split into respective clusters of psychological processes according to their designation within the LIWC dictionary: style measures, complexity, function words, emotional processes, cognitive processes, perceptual processes, biological processes, motivational

processes, temporal processes, spatial-relational processes, personal processes, and utterances (see Table 4.3).

Psychometrics of the Mental Profile Map

Calculation of distance metrics

Psychometrics were calculated separately for each author in order to verify that results were generalizable across individuals and not idiosyncratic or unique to a single playwright. For the plays of Aphra Behn, only those plays designated as verified and legitimate Behn plays, including adaptations (i.e., those listed in Table 1), were included in the psychometric analysis. William Rowley was excluded from the psychometric analysis (and all subsequent MPM procedures) due to the fact that only 1 play by this author survives, precluding the ability to establish reliability over time.

Following the clustering of psychological processes, the center point of each psychological cluster was calculated separately for each author using a robust bootstrapping method (iteration $N = 1000$). For example, the "Complexity" cluster of psychological processes has a 2-dimensional center point for Lewis Theobald: the average of his "Analytic" scores for all of his plays is one dimension, and the average of his "Sixltr" scores for a second dimension. Similarly, the "Social" cluster has a 5-dimensional center point, and so on. Theobald's center points for each process were uniquely derived from his own works, as were the center points for Shakespeare, Behn, and the remaining authors derived from their respective works. After calculating the center point for each individual for each psychological process, the distance of each play from the center was calculated using the Mahalanobis distance statistic. This effectively resulted in 13 separate Mahalanobis distance metrics per written work, each existing within an author's mental profile map.

The bootstrapped Mahalanobis distance procedure resulted in a separate "distance from center" score for each psychological process for each play. These scores were then squared and converted to 0-100 scores using a chi-square estimation, resulting in a series of 13 distance measures for each play. Using this 0-100 scale, a theoretical score of 100 would reflect plays sitting perfectly at the psychological center of the map, and a score of 0 would reflect plays that fall extremely far from center. For example, if a play had a score of 90.5 for the perceptual processes measure, it would be relatively near the center of the author's overall perceptual processes map. A low score, such as 10, would be extremely far away from an author's perceptual processes center.

Note that the Mahalanobis metrics are, in a way, silent about directionality of deviations. Simply put, if a single distance measure (e.g., motivational processes) is extreme for a given play, the core interpretation is simply that this play is functionally *different* from the other works in this respect. The distance measure itself does not reveal, for example, that a certain play was deeply laden with (or bereft of) words related to motivation processes – the measure only demonstrates that a certain play appears to be askew given the assumption that all plays were authored by the same individual. Instead, a score that shows great distance from the psychological center tells us that the *composition* of motivational processes for the play in question is markedly different from the other works of an author. While not inherently providing directionality information, these scores can subsequently be decomposed into meaningful interpretations, such as "While Shakespeare's texts generally have high 'affiliation' words and low 'power' words, the pattern reverses for *Romeo and Juliet*, suggesting a radically different balance of motivational processes.". Decomposition and interpretation of MPM analyses are later described.

Internal consistency of distance metrics

After calculating the "distance from center" scores for each psychological process for each play for each author, Cronbach's alpha was used to determine whether distance scores were consistent with each other. For example, if a play is far from the psychological center of Shakespeare's collected works on cognitive processes, is the same play generally an outlier across *all* psychological processes, or are these distance metrics relatively isolated?

Cronbach's alpha results are presented in Table 4.4. The internal consistency analysis did, in fact, find that the distances of each psychological process for each play formed a coherent single factor for all authors in the dataset. In other words, when a given play was further away from an author's center along one set of psychological processes, it was generally further away from the center along all other psychological processes as well. These results were universally true and not isolated to a specific author or subset of authors. These results are congruent with other work on the internal consistency of psychological measures of language (see Pennebaker, Boyd, Jordan, & Blackburn, 2015) and demonstrate that language-based measures of psychological processes do, in fact, vary in unison from the center.

Author	Cronbach's Alpha of Distance Measures
Behn	.66
Fletcher	.54
Marlowe	.64
Shakespeare	.52
Theobald	.70
Dekker	.66

Table 4.4. Internal consistency of distance measures for each author.

The description of the MPM steps performed up to this point is, admittedly, rather intricate. As an illustrative example of the procedures described thus far, Table 4.5 presents descriptive statistics and inter-item correlations for all distance measures for Behn's plays. In effect, for each of the plays of Apha Behn, a separate Mahalanobis distance score was calculated for each of the 13 broad psychological processes captured by LIWC. Put another way, all 16 of Behn's plays were each assigned 13 separate "distance from center" scores, one for each broad cluster of psychological processes. Across all 13 psychological processes, Behn's plays were, on average, a modest distance away from the center points, with all scores hovering closely around 50. Furthermore, the distance scores for each of these measures were strongly intercorrelated, meaning that variation across all psychological processes were relatively harmonized with each other. For example, in Behn's plays that showed drastic deviation from (or adherence to) the center in terms of emotional processes, so too did they show considerable deviation in biological and motivational processes, and so on.

The fact that the psychological distance scores all converge on similar information facilitates two useful procedures in moving forward. First, given that all distance measures appear to be reflecting a similar phenomenon (distance from the "psychological center" of the map), one is able to collapse all distance scores into a single psychological processes down to a single score.

Distance Measures	Mean (SD)	1	2	3	4	5	6	7	8	9	10	11	12
1: Style	46.24 (26.87)	–											
2: Complexity	48.05 (30.81)	-0.152	–										
3: Function Words	45.34 (14.32)	0.41	0.281	–									
4: Emotional	49.08 (28.42)	0.337	-0.234	-0.143	–								
5: Social	50.29 (28.67)	0.21	-0.2	-0.324	0.23	–							
6: Cognitive	47.14 (27.31)	0.019	0.179	0.439	-0.388	-0.317	–						
7: Perceptual	50.36 (28.47)	0.163	0.301	0.037	0.188	0.626	-0.343	–					
8: Biological	54.16 (29.81)	0.522	0.006	0.057	0.467	0.503	-0.321	0.567	–				
9: Motivational	53.49 (31.59)	0.497	-0.474	-0.141	0.618	0.363	-0.095	0.193	0.391	–			
10: Temporal	47.03 (30.58)	0.297	0.38	0.232	-0.176	-0.196	0.312	-0.242	-0.116	-0.178	–		
11: Relational	47.01 (26.66)	0.641	-0.118	0.286	0.277	0.17	-0.177	0.174	0.448	0.052	-0.071	–	
12: Personal	49.87 (29.72)	0.542	0.072	0.198	0.545	0.154	-0.269	0.194	0.714	0.276	-0.035	0.515	–
13: Utterances	50.41 (29.39)	-0.221	0.555	0.122	0.03	-0.224	0.118	0.088	0.209	-0.271	-0.016	-0.219	0.35

Table 4.5. Summary statistics and inter-item correlations for all distance measures calculated for the verified plays of Aphra Behn (i.e., excluding questioned plays).

Note: The resulting Cronbach's alpha for all distance measures ($\alpha = .66$) suggests sufficient intercorrelation to constitute collapsing the distance measures into a single, meaningful psychological metric.

The median of all distance scores was calculated for each play[4], resulting in a single 0-to-100 score that signified the strength for the case of an individual's authorship for any given work (e.g., Behn's M = 48.56, SD = 16.41). This final score is essentially a shorthand metric that tells us that whether a play is generally far away from the "psychological center" of the map across a battery of psychological processes, in which case it would merit further investigation/explanation.

Second, one can meaningfully reduce the number of dimensions from 13 psychological distance scores to 2 using multidimensional scaling (e.g., a principal components analysis). Reducing the dimensionality is primarily useful in this case to create a visualization that roughly represents how far away from the center each play rests. This allows for the consumer of these results to more easily and intuitively understand the concept of the MPM method, providing a visual reference for discussion, analysis, and interpretation of results. Given the nature of single-candidate authorship attribution problems and the inclusion of a relatively small sample of works, a mental profile map allows us to get a sense of how far away a questioned work sits from the overall psychological center. Questioned works that show relatively low MPM scores should then be subjected to further scrutiny and questioning.

MPM visualizations for all authors included in MPM analyses are presented in Figure 4.1. Note that while each author may have one or two plays that stray from the broader cluster of their psychological maps, all plots show a central "gravity" that visually represents a psychological center across all mental processes. In other words, all written works by all playwrights exhibit a tendency to radiate out from the psychological center of their

[4] In this case, the median is preferred over the mean to prevent undue influence of a single psychological process outlier. Much like the unmasking process, drastic variations along one or two psychological processes may be superficial (e.g., different themes between works) and not reflect meaningful psychological differences (or similarities). Use of the median over the mean helps to reduce excessive influence of superficial extremities. However, in this case, mean and median distance scores showed a strong correlation (R = +0.96).

respective map, demonstrating that each author's works deviate from their own, internal norm across an entire psychological spectrum rather than single psychological processes. The center point (0,0) for each subgraph of Figure 4.1 is the approximate "psychological center" for each playwright when collapsing across all 13 psychological process categories. Note that the projection of distance measures down to 2 dimensions does result in some distortion, and this graph should be interpreted as an approximation of the "true" location of each play. In other words, some plays may actually be somewhat closer to (or farther from) the center of each map than what the image may suggest.

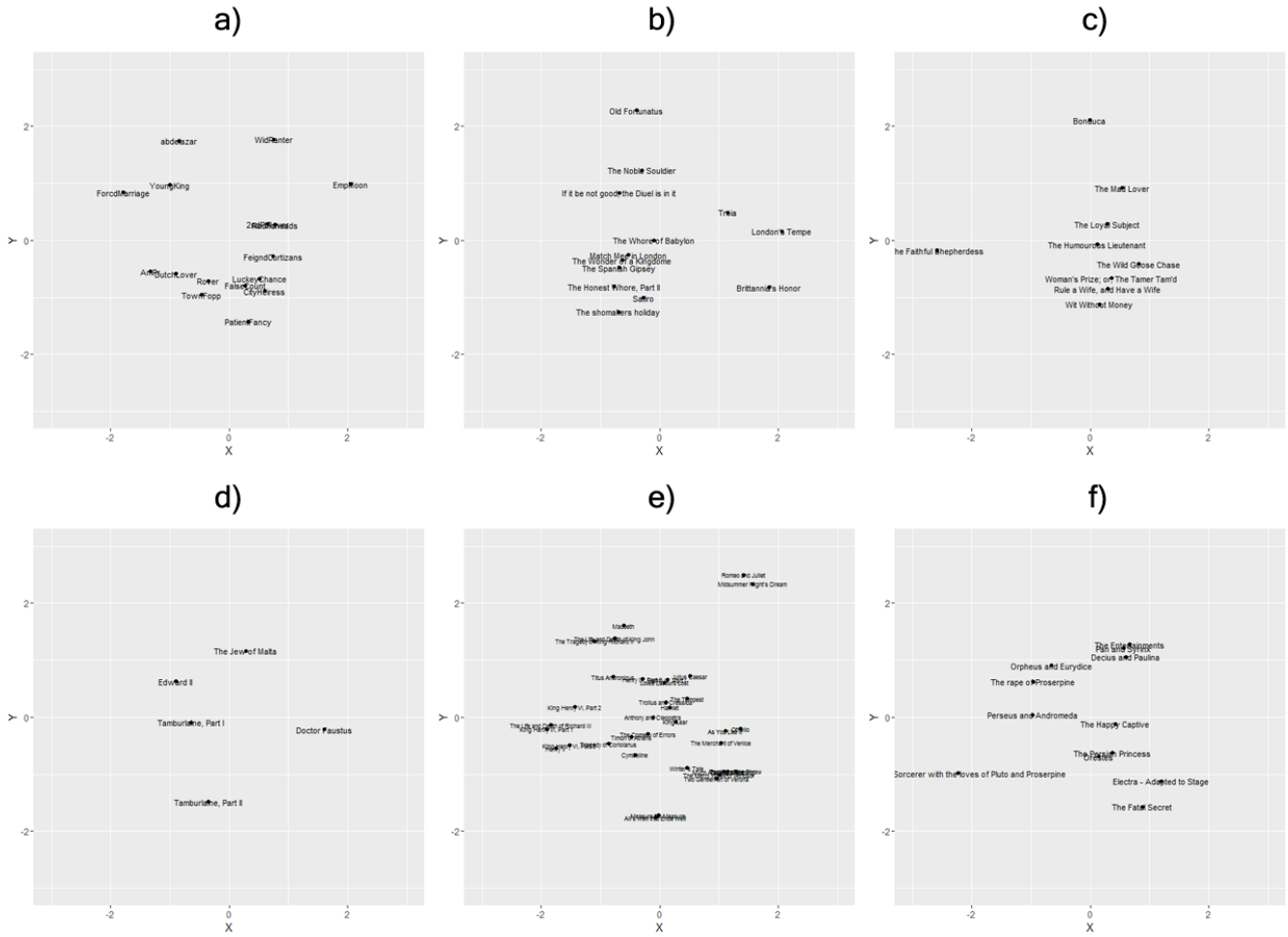

Figure 4.1: Visualized mental profile maps of 6 playwrights: a) Aphra Behn, b) Thomas Dekker, c) John Fletcher, d) Christopher Marlowe, e) William Shakespeare, and f) Lewis Theobald.

Testing the Mental Profile Map Approach for Authorship Tasks

In the case of an authorship attribution task, we can adopt the MPM approach by creating a map using all of an author's known works and, in turn, also including each work of questionable origins. By doing so, one is able to generate scores for each individual questioned work and make judgments regarding how well the questioned works fit into the larger picture. This is the approach that was adopted for all MPM analyses reported below.

As an initial test of the utility of the MPM approach in authorship attribution tasks, it is useful to first examine its performance in cases where all works are of known authorship. This can be done by performing similar procedures to those described above, albeit with some "bogus" insertions of comparable works by other known authors. It is possible, for example, to insert a play by William Shakespeare into Aphra Behn's map. By doing this using the MPM procedures described above, we are essentially operating under the known bogus assumption that Behn actually authored Shakespeare's play. Operating under this bogus assumption, we should be able to spot the false insertion due to its drastic pulling away from the MPM center, both numerically and visually.

For this test analysis, three plays were chosen by a random number generator from the supplemental playwright corpus. These 3 random plays included *The Whore of Babylon* by Thomas Dekker, *The Fatal Secret* by Lewis Theobald, and *Julius Caesar* by William Shakespeare. Each play was, in turn, inserted into the corpus of verified plays by Aphra Behn – the MPM procedures described in the preceding section were then performed. This resulted in 3 separate "maps", one for each run of the MPM procedure with the inclusion of each bogus play insertion. Numeric results are presented in Table 4.6; visualizations are presented in Figure 4.2.

Author	Title	Mental Profile Map Score
Behn	The Lucky Chance	76.59
Behn	Sir Patient Fancy	69.41
Behn	The Young King	68.98
Behn	The Feigned Courtesans	61.26
Behn	The False Count	60.56
Behn	The Town Fop	60.56
Behn	The Dutch Lover	58.88
Behn	The Rover, Part I	53.59
Behn	The City Heiress	50.00
Behn	The Roundheads	46.36
Behn	The Forc'd Marriage	45.76
Behn	The Rover, Part II	42.13
Behn	The Amorous Prince	41.02
Behn	Abdelazer	33.98
Behn	The Emperor of the	24.43
Shakespear	Julius Caesar	24.20
Theobald	The Fatal Secret	21.11
Dekker	The Whore of Babylon	18.09
Behn	The Widow Ranter	17.64

Table 4.6: Results of the MPM analysis with bogus play insertions.

Note: MPM scores for plays marked as "Behn" authorship are averaged across each MPM analysis (correlation with Behn-only analysis: r = .95). Works highlighted in yellow are those that were artificially inserted as bogus Behn plays.

When interpreting Figure 4.2, consider that the blue diamond denotes the psychological center of each map. Bogus plays are circled in red and stand out considerably from the rest of the map in each case. Bogus plays include Dekker's *The Whore of Babylon* (left), Shakespeare's *Julius Caesar* (middle), and Theobald's *The Fatal Secret* (right).

Results from the "bogus play" MPM analysis were extremely promising. All 3 bogus plays that were inserted are markedly distinct from the rest of Behn's map. Numerically, all 3 insertions scored extremely low for fit in Behn's corpus, with only one

verified Behn play (*The Widow Ranter*) scoring as a worse fit than all three; the outstanding Behn play is discussed later. Visually, each bogus play was also quite distinct, falling well outside of Behn's relatively tight ring of verified works. These results strongly suggest that this method is useful for identifying gross psychological departures from an author's norm or, in this case, works that show a psychological signature of someone other than an author in question.

Results: Mental Profile Map Analysis of Questioned Plays

The above analyses show a strong potential for the use of MPM in authorship attribution tasks. An analysis of the questioned Aphra Behn plays was thus performed in a manner parallel to that described above. Rather than inserting bogus plays into the mental profile map, however, each questioned work was inserted in turn. Numeric results from the MPM analyses are shown in Table 4.7.

In this case, an analysis of the MPM scores would suggest that two questioned plays, *The Younger Brother* and *The Revenge*, show psychological patterns that are fairly typical of Behn's corpus of plays altogether. Indeed, both of these questioned plays fare above average in their comparison to Behn's works of verified authorship. Conversely, the remaining three questioned plays show rather low MPM scores, suggesting a

Figure 4.2: Visualized mental profile maps of Aphra Behn when including bogus plays by other playwrights.

Author	Title	Grand MPM Score (Median)	1	2	3	4	5	6	7	8	9	10	11	12	13
Behn	The Lucky Chance	72.02	69.42	96.94	68.38	14.84	50.54	72.02	75.64	75.21	3.04	94.39	81.37	62.60	87.84
Behn	The Young King	70.56	83.71	44.34	73.98	70.56	47.75	19.60	90.49	81.69	69.88	33.71	85.94	75.14	12.98
Questioned	The Younger Brother	67.60	67.60	68.81	29.27	54.76	27.29	88.03	33.52	53.14	71.21	77.16	10.11	88.00	79.82
Behn	Sir Patient Fancy	66.26	94.83	29.76	29.22	66.26	73.22	29.03	39.74	77.54	83.89	92.20	70.82	47.62	5.38
Behn	The Feigned Courtesans	62.54	41.50	60.35	36.47	93.81	62.54	21.89	48.67	65.31	85.75	84.21	30.48	84.60	66.79
Behn	The False Count	58.94	36.57	93.89	42.78	68.39	17.45	69.25	67.80	57.58	76.97	30.20	28.58	58.94	89.54
Behn	The Town Fop	56.17	56.17	72.93	65.45	23.58	50.61	75.45	39.49	76.23	54.98	45.95	28.91	72.18	88.46
Behn	The Rover, Part I	54.98	91.05	14.78	54.98	68.63	32.55	73.98	13.39	35.02	85.04	41.39	90.57	80.09	40.48
Behn	The City Heiress	50.75	50.75	49.00	36.70	68.04	82.15	18.78	96.26	68.83	75.59	38.96	13.39	45.18	77.44
Questioned	The Revenge	49.78	92.11	90.84	28.70	4.30	7.51	31.98	68.13	58.54	59.78	70.88	45.73	32.47	49.78
Behn	The Forc'd Marriage	45.99	45.99	13.37	25.63	31.94	88.31	11.08	68.95	84.46	65.84	9.33	56.16	80.95	29.76
Behn	The Roundheads	44.51	44.51	16.89	55.40	49.21	53.77	85.26	35.96	27.67	78.52	59.36	25.46	7.69	3.03
Behn	The Amorous Prince	42.31	29.00	19.26	33.14	42.31	88.81	70.89	73.44	70.77	79.00	13.71	44.97	17.98	38.26
Behn	The Rover, Part II	40.83	22.85	64.91	40.83	40.42	3.79	27.25	31.95	54.26	20.72	45.90	52.28	41.12	59.97
Behn	The Dutch Lover	40.31	32.16	31.30	39.82	98.54	60.88	29.40	40.31	81.61	29.74	12.27	69.63	87.87	78.13
Behn	Abdelazer	35.53	8.56	92.99	36.36	20.38	74.22	35.53	71.12	6.14	5.37	35.97	45.08	9.91	31.47
Questioned	The Counterfeit Bridegroom	34.60	17.31	86.96	30.93	34.60	2.51	43.92	42.09	87.44	35.15	10.21	5.04	85.49	29.15
Questioned	The Woman Turned Bully	33.58	3.61	69.86	31.66	18.59	59.72	20.98	33.58	55.48	44.41	5.25	6.53	52.57	96.00
Questioned	The Debauchee	32.34	37.53	32.34	29.25	2.46	29.24	45.32	8.71	26.37	39.47	20.70	49.54	39.25	66.22
Behn	The Widow Ranter	20.03	20.03	61.74	46.74	2.12	4.40	86.16	1.48	1.34	3.12	99.67	4.71	20.69	48.53
Behn	The Emperor of the Moon	15.23	12.70	6.35	39.41	26.21	13.65	28.59	11.05	2.89	38.38	15.23	23.70	5.30	48.47

Table 4.7: Results of the MPM analysis for Behn and questioned works.
MPM scores for plays marked with "Behn" authorship are the result of the MPM analysis that included only verified Behn plays. Works highlighted in yellow are those of questioned authorship. Higher Grand MPM scores are indicative of plays with a general low distance from center (i.e., a better fit with Behn's mental profile map).

Figure 4.3: Visualization of Behn's mental profile map when including The Revenge (left), a play that shows a strong MPM score, and The Debauchee (right), a play that shows a weak MPM score.

moderate-to-high psychological distance from verified Behn works. The patterns are particularly striking when considering visualizations of the questioned plays' locations within the mental profile maps (examples presented in Figure 4.3).

When interpreting the map visualizations note that, on average, the questioned play *The Revenge* appears to be highly prototypical of Behn's mental profile, falling close to the center of the map. Plays with rather low MPM scores, such as *The Debauchee*, show rather distant positions from the center and break from the "orbiting" placement of most other plays in terms of its psychological profile.

Taking both the MPM scores and visualizations together, these analyses suggest that 3 of the 5 questioned plays – *The Debauchee, The Woman Turned Bully,* and *The Counterfeit Bridegroom* – remain highly suspect regarding Behn's authorship. These

results are perfectly convergent with those provided by the earlier "unmasking" method. Given the high reliability of the language-based measures used for the current analysis, as well as the impossibility of achieving the "just right" balance between nearly 80 language dimensions without the aid of computerized systems, these results are extremely unlikely to occur by chance.

Mental Profile Map Decomposition

Decomposition of questioned plays

A primary benefit of the MPM method over more opaque machine learning methods is the ability to decompose the results into interpretable, meaningful units of analysis. Rather than receiving a hard probability score as the final result, it is possible with the MPM method to peer under the hood and look for reasons as to *why* a given text appears to be a poor fit. This can be particularly useful for looking for clues as to a work's true author (if identified as a poor fit), or better understanding a drastic variation (e.g., a recent traumatic event or other upheavals) when authorship is certain.

The results presented above in Table 4.7 include each of the 13 distance metrics for each play, allowing us to manually inspect those psychological processes that appear to be driving the effects above. For all three questioned plays that show very low MPM scores (i.e., *The Debauchee, The Woman Turned Bully,* and *The Counterfeit Bridegroom*), it is clear that they generally show below-average MPM scores, suggesting a generally great distance from Behn's psychological center along most psychological processes. Table 8 highlights those psychological processes along which the 3 low-scoring questioned plays are most discrepant (i.e., MPM scores <= 20).

Play Title	Highly Discrepant Psychological Processes
The Counterfeit Bridegroom	Style, Social, Temporal, Relational
The Woman Turned Bully	Style, Emotional, Motivational, Temporal
The Debauchee	Affect, Perceptual

Table 4.8. Numeric results for psychological processes that showed particularly great distance from center for the 3 questioned plays with poor support (i.e., MPM scores <= 20) for Behn's authorship.

Once broad psychological discrepancies have been identified, it is possible to then return to the raw data to examine the specific psychological constructs that are driving the differences between the questionable plays and Behn's overarching mental profile. In doing so, one can reference the raw scores underlying the MPM distance scores (available from the corresponding author by request) to look for extremities.

In a manual analysis, it is possible to see that *The Counterfeit Bridegroom* scores extremely high on stylistic measures such as clout and authenticity, and overall social words, family words and friend words. Additionally, this play showed extremely low past-focus and high future-focus within the dataset, and a generally high score on time words from the relational processes cluster. This combination of extremities suggests an author with several discernible psychological characteristics: an extremely socially focused individual with strong social standing, and likely an individual who is also highly goal-directed in their day-to-day behaviors, as evidenced by the high use of future and time concepts.

Similarly, the play *The Woman Turned Bully* showed a number of extreme psychological differences from the profile extracted from Behn's other works. In decomposing the psychological processes of this play, several major differences were apparent. *The Woman Turned Bully* exhibited extremely high authenticity scores, very low

affect words (including both positive and negative emotions, and negative emotion subtypes), and low past-focus and high present-focus. Additionally, this play exhibited large differences from Behn's profile in both reward words (very high) and risk words (very low). Taken together, the psychological profile of this play suggests an author who is extremely impulsive, focused on the "here and now", low in self-monitoring, and possesses a strong drive for reward at the cost of risk sensitivity.

Of the three questioned plays, *The Debauchee* is perhaps the most generally different from Behn's mental profile, yet in very few extreme ways. In emotional terms, for example, this play included extremely high use of general negative emotion words (but low use of specific negative emotion words, such as sadness or anger) and low use of positive emotion words. The other extremities for this play occurred in the domain of perceptual processes, with this play exhibiting extremely low use of perceptual words (e.g., "see" words and "feel" words), but high use of words related to sound. This small combination of extremities is generally difficult to interpret; the broader pattern of a more generalized distance from Behn's profile may instead simply suggest a person whose psychology is fundamentally different from Behn in most ways.

Decomposition of Behn's outlying plays

In the course of the MPM analyses, two additional plays that were included as accepted works by Behn also demonstrated an extreme divergence from the psychological center of Behn's map. Like the questioned play *The Debauchee*, The *Emperor of the Moon* exhibited a broad, generalized difference from the mental profile of Behn, with no specific clusters of psychological processes appearing to be particularly outstanding (i.e., MPM score <= 20); instead, virtually all processes were outstanding. *The Widow Ranter* was highly discrepant in both the "bogus insertion" analysis described earlier as well as the

MPM analysis of questioned plays. *The Widow Ranter* shows a more unique pattern: several of the psychological processes (complexity, cognitive, and temporal) appear to be a very close fit with Behn's profile, whereas the others have varying degrees of distance between her psychological center.

Unlike the results provided by the "unmasking" method, which might only suggest that these 2 plays would be difficult to classify, we are able to look for reasons as to why such plays may vary so drastically. Given the historical context, it is difficult to conclusively determine the forensic history of these two accepted plays. One possible explanation for the extreme positioning of these 2 plays would revolve around the nature of collaboration amongst playwrights during the time of Behn. For example, it is generally accepted that notions of authorship and collaboration were rather different than those of today, and there is extensive evidence that uncredited joint authorship was commonplace within the King's Company and Duke's Company (see Bentley, 1986; Vickers, 2004), both of which were groups for which Behn worked. As such, it is possible that these plays were either only partially authored by Behn, or perhaps heavily revised by other authors.

Additionally, both *The Widow Ranter* and *The Emperor of the Moon* were likely written near the time of Behn's death in 1689 (O'Donnell, 2004; Korte, 2015). In her final years, Behn's health was ailing and the nature of her work saw a shift, including various other types of prose and translations. The fact that Behn began to suffer from poor health may have been coupled with the accompanying psychological shifts (e.g., Shaffer, 2000) and could potentially explain the drastic change in the mental profile of these two plays relative to the other works of Behn.

DISCUSSION

The current study used the works of several playwrights to introduce a new, sophisticated authorship attribution methods for cases of single-candidate attribution tasks. Results from this new "Mental Profile Mapping" method were also compared with a powerful attribution method from the computational sciences known as "unmasking". Throughout the course of the study, several goals were achieved: the development of a new method for single-candidate problems, the shedding of light on a specific authorship question, and the establishment of psychometric possibilities in the realm of high-dimensional psychological profiling.

Mental Profile Mapping Method

The current study demonstrated the underlying methods and utility of a new method, MPM. The MPM approach to authorship attribution possesses a number of benefits over other methodologies in the authorship attribution space. Foremost among these benefits is that it is, to the author's knowledge, the only existing "pure" method for single-candidate authorship attribution methods. Unlike other methods that require the introduction of additional data from other sources, either in the form of imposters or comparators for modeling, the MPM method requires "ground truth" text from only a single source. In cases where comparable texts from other authors are unavailable, this feature of the MPM method is particularly valuable.

Results from all analyses in the MPM framework were particularly strong. The psychometric assessment of the MPM as a methodology revealed that its underpinnings do, in fact, form a coherent construct that suggests language samples vary not just in single, isolated ways from a person's psychological center but, instead, across all included language-based measures of psychology in unison. Additionally, "ground truth" tests that

included known bogus insertions performed extremely well – the MPM method was able to capture false Behn plays almost perfectly.

As demonstrated in the current study, the MPM method need not exist or be performed in isolation. In cases where additional "supplemental" texts can be made available for the purposes of establishing baseline functions, several advanced machine learning methods may be used in conjunction with the MPM approach to strengthen an inquirer's confidence in the results. For example, if multiple, radically different attribution methods converge on similar results, as occurred in the current study, increased confidence can be placed in the outcome. Additionally, features of the MPM method, such as the discrete psychological process distance scores, may be useful for inclusion in other authorship attribution frameworks.

Finally, a featured benefit of the MPM approach over other authorship attribution methods is that it is fundamentally a *psychological* method of authorship attribution. Insofar as psychological information can be extracted from language data, an individual performing the MPM method is able to more deeply and thoroughly examine an attribution problem by combining the results with data from other sources. By decomposing the MPM distance scores, researchers are able to identify cases of questionable origin but, also, extract a psychological profile from questioned texts. This ability may be of particular value in legal and forensic settings, where various forms of evidence and information must be considered together in order to render decisions. For example, if the MPM of an authorial suspect closely aligns with behavioral outcomes (e.g., never late to work, always polite) or personality reports from family and friends (e.g., conscientious, agreeable), a questioned work that shows radically different psychological properties (e.g., hostile and impulsive) is likely to not only show statistical differences from a candidate's MPM, but can point to the psychological profile of the true author.

The Plays of Aphra Behn

In the current study, the MPM procedure was put the test using the works of Aphra Behn, a prolific female playwright of the 17th century. Across 2 highly distinct attribution methodologies, Behn's unique psychological fingerprint was discernible in her work. This remained true with her original works and, additionally, Behn's adaptations were also clearly imprinted with her unique psychological composition.

Of primary focus in the current test were 5 plays of questioned origins. The unmasking analysis and the MPM analysis converged to identify 2 of the 5 questioned plays, *The Revenge* and *The Younger Brother*, as showing a high likelihood of Behn's authorship. The remaining 3 plays, *The Debauchee, The Woman Turned Bully,* and *The Counterfeit Bridegroom,* exhibited an extremely poor fit for Behn's mental profile across both attribution methodologies. Additionally, the MPM analysis provided results suggesting likely psychological traits of the 3 questioned plays' authors. *The Woman Turned Bully* bears the signature of a highly impulsive person with poor self-monitoring abilities. *The Counterfeit Bridegroom* exhibits language patterns that are commonly associated with individuals of particularly high social standing and a strong, goal-oriented mindset. *The Debauchee*, unlike the other 2 plays of questionable origin, did not show any particularly outstanding mental profile – rather, the embedded psychological traits appeared to be, quite simply, generally different from those of Behn.

Like all automated authorship attribution studies, care should be taken in interpreting the results of this authorship test. In the world of authorship attribution, particularly with historical data of uncertain origins, results are never a "sure thing" and must be interpreted in the light of converging evidence. In other words, the results of these analyses cannot conclusively *prove* that Behn did not have an authorial hand in *The Debauchee, The Woman Turned Bully,* and *The Counterfeit Bridegroom*. However, the

results of both the unmasking and MPM procedures provide a strong impetus for deeper examination from domain experts researching the area of Aphra Behn's life and work. At the very least, the 3 weak-evidence plays merit some explanation for their divergence from Behn's verified works on the mental profile map – an explanation is further merited by the results of the unmasking method, which supports the conclusions of the MPM approach.

Limitations and Future Directions

The current study does possess some limitations that should be taken into consideration. Functionally speaking, the current study includes a small sample size; only 6 to 7 authors were included for all of the analyses performed in this work. It is possible that with a greater number of authors included in the current sample, the results of all analyses may shift to favor another conclusion. Importantly, however, the unmasking method has been extensively validated and tested in previous work (Koppel et al., 2007). In other words, the unmasking method is already proven and established as a valid and powerful form of addressing authorship attribution questions. The convergence of the MPM results with the unmasking method is extremely promising. Nevertheless, future work may benefit from more extensive testing on broader samples.

Additionally, the current study was performed in a rather constrained context. In practical terms, a demonstration of the MPM method on Elizabethan, Stuart period, and Restoration era, playwrights may not extend to texts of other eras or genres. Additionally, all texts used in the current analysis were of particularly healthy length. For example, in the Behn/questioned work MPM analyses, the average word count was nearly 26,000 words per play. The degree to which the MPM method would be applicable to shorter texts, such as e-mails, short letters, or social media updates, is unclear.

In spite of the constrained context, there is no reason to suspect that the procedures used within the MPM method would not extend to domains outside of the current tests. Indeed, most authorship attribution tasks are initially tested on long-form texts such as novels, yet are still viewed as applicable to other forms of language samples, assuming that all texts in a given sample are of generally comparable genres. Additionally, the language-based measures used in the current analyses have been extensively validated for the purpose of extracting psychological information from language, often including extremely short texts such as tweets (e.g., De Choudhury, Gamon, Counts, & Horvitz, 2013; Sylwester & Purver, 2015).

Nevertheless, future work with the MPM method should focus on an expansion outside of the current context. The most obvious applications for the MPM method are in both legal and forensic contexts wherein a reconstruction of historical events, such as determining a document's origins, may be absolutely vital for rendering verdicts of guilt or innocence. Further support for the conclusions of the MPM method can be provided in the context of machine learning frameworks as well; the distance metrics generated as a part of the MPM procedures will likely prove useful in the context of other machine learning and authorship attribution frameworks.

Crucially, the underlying concept of the MPM approach may possess extended value outside of forensic applications. In the modern world, technology has opened up diverse and highly complex possibilities for new assessment methods, such as the collection of rich behavioral data from smartphone technologies (e.g., Harari et al., in press). As interest grows in the use of rich idiographic data in the fields of mental health and medicine (e.g., Bakker et al., 2016), new methods of quantifying an individual's psychological variations over time will be required. The MPM and methods like it could, for example, be put to meaningful use in clinical settings for patients with psychotic or

mood disorders. Such an approach may allow mental health providers to more accurately monitor patients' day-to-day psychological variations and potentially facilitate faster detection of extreme, generalized psychological variations that could be diagnostic of problematic episodes. Such possibilities currently remain within the scope of future research in psychology and the computational sciences.

Conclusions

Mental Profile Mapping is an early first step in realizing the possibilities of pairing advanced statistical modeling procedures with interpretable, actionable psychological insights. Additionally, the current work highlights the future promise of better understanding the individual as a high-dimensional composite or bundle of psychological processes. As psychological and computational forensic techniques continue to advance, there will be an increasing number of opportunities to create meaningful combinations of methods from the two disciplines. Future work in the areas of psychological and computational forensics will likely benefit from increased collaboration and cross-communication. As new techniques are needed to address increasingly complex and nuanced problems in each field, the adoption of techniques from both areas of study will help to generate more comprehensive, rigorous, and meaningful insights into the human condition.

Chapter 5: General Discussion

This dissertation has presented 3 separate lines of research that demonstrate both the current ability of researchers to extract high-dimensional psychological information using language analysis techniques and the utility of doing so. In 2 authorship attribution studies and 1 study of human values, various language analysis techniques paired with a diverse array of statistical and machine learning algorithms were used to perform tasks as diverse as identifying an individual based on a psychological "fingerprint" to exploring the deep properties of high-dimensional linguistic consistency within an individual over time.

In Chapter 2, psychological fingerprinting was used to not only differentiate multiple authors but, additionally, to check for convergence between authorial psychological profiles and historical observer reports. In this case, the more high-powered machine learning algorithms (e.g., SMO support vector machines) were able to take language-based psychological metrics to differentiate authors with extremely high accuracy. Additionally, more traditional statistical approaches (i.e., linear discriminant analysis) were shown to provide a generally interpretable model of unique psychological attributes for each author, allowing for a merging of traditional goals from the field of psychology (i.e., interpretation and understanding) with computational statistical approaches (i.e., predictive accuracy and error minimization).

Chapter 3 extended the concept of high-dimensional measurement of psychological phenomena to the domain of human values by adopting a more traditional psychological study design. By using a bottom-up language analytic method, it was shown that the use of open-ended language samples could be used to more powerfully and diversely predict real-life, every day human behaviors than what is often the case

with a widely-used "gold standard" self-report questionnaire. Additionally, by adopting a naturalistic observational approach in Project 2, it was found that language far outperformed the traditional self-report questionnaire, again both in terms of breadth and interpretability.

Finally, Chapter 4 returned to the concept of authorship attribution to address the difficult problem space of single-candidate authorship attribution. In Chapter 4, psychological measures of language were paired with a sophisticated meta-learning procedure (i.e., "unmasking") to successfully differentiate authors as a function of their prediction degradation curves as language features were iteratively removed. Additionally, the concept of mental profile mapping (MPM) was introduced as a new approach to single-candidate authorship attribution. Results from the MPM approach converged closely with the more complex attribution technique, yet affords researchers the ability to sequentially decompose distance metrics in order to describe outliers in psychological terms.

CONCLUSIONS

Overall, the 3 sets of studies included in this dissertation can be taken to demonstrate the usefulness of extracting psychological information from language as a high-dimensional measurement technique. While the research on value measurement via language shows promise as an alternative to self-report measures, affording stronger predictive and ecological validity than traditional approaches, the authorship attribution studies demonstrate the importance of such techniques wherein alternative, objective psychometric methodologies do not exist. In all cases, by using language as a means for extracting a high number of measures along multiple psychological dimensions, it was

possible to take widely available naturally occurring data (i.e., language) and extract psychological information in an extremely broad and accurate fashion.

As higher quantities and qualities of data become pervasive and accessible to researchers via social and other digital media, the application of new high-dimensional assessment techniques will be able to not only proliferate more rapidly (as they already have), but will serve as complementary approaches to psychological research. By combining high-dimensional psychometrics from language with measures extracted from other modalities (e.g., still imagery, video, audio), more refined approaches to the measurement of the individual will undoubtedly emerge, providing researchers with new, powerful tools to measure and better understand the human condition.

FINAL NOTES

The studies included in this dissertation were all conducted in conformance standard ethical guidelines. The author of this dissertation (Ryan L. Boyd) was the primary researcher for the published studies and was the principle individual involved in the data analyses and writing. The final study presented in this dissertation has not yet been published, but was exclusively conducted and written by the author of this dissertation. All content in the current dissertation is either newly-written content that is unique to this document or taken from pre-published manuscripts of works that were later altered in the publication process.

Appendix

Plays by Shakespeare	Plays by Fletcher	Plays by Theobald
A Midsummer Night's Dream	Bonduca	Decius and Paulina
All's Well that Ends Well	Rule a Wife, and Have a Wife	Electra
Antony and Cleopatra	The Faithful Shepherdess	Harlequin a Sorcerer
As You Like It	The Humourous Lieutenant	Orestes
Coriolanus	The Loyal Subject	Orpheus and Eurydice
Cymbeline	The Mad Lover	Pan and Syrinx
Hamlet	The Wild Goose Chase	Perseus and Andromeda
Henry IV, Part 2	The Woman's Prize	The Fatal Secret
Henry IV, Part I	Wit Without Money	The Happy Captive
Henry V		The Lady's Triumph
Henry VI, Part 1		The Persian Princess
Henry VI, Part 2		The Rape of Proserpine
Henry VI, Part 3		
Julius Caesar		
King John		
King Lear		
Love’s Labours Lost		
Macbeth		
Measure for Measure		
Much Ado About Nothing		
Othello		
Richard II		
Richard III		
Romeo and Juliet		
The Comedy of Errors		
The Merchant of Venice		
The Merry Wives of Windsor		
The Taming of the Shrew		
The Tempest		
The Two Gentlemen of Verona		
The Winter's Tale		
Troilus and Cressida		
Twelfth Night		

References

Aronson, E. (2004). *The social animal*. New York, New York, USA: Worth Publishers, 9th edition.

Back, M. D., Küfner, A. P., & Egloff, B. (2010). The emotional timeline of September 11, 2001. *Psychological Science, 21*(10), 1417-1419.

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: Review and evidence-based recommendations for future developments. *JMIR Mental Health, 3*(1), e7. http://doi.org/10.2196/mental.4984

Ball-Rokeach, S., Rokeach, M., & Grube, J. W. (1984). *The great American values test: Influencing behavior and belief through television*. New York, New York, USA: Free Press.

Barkan, L. (2001). What did Shakespeare read? In de Grazia, M., & Wells, S. (Eds.), *The Cambridge companion to Shakespeare* (31-48). Cambridge: Cambridge University Press.

Bentley, G. E. (1986). *Profession of dramatists in Shakespeare's time, 1590-1642*. Princeton, NJ: Princeton University Press.

Beukeboom, C. J., Tanis, M., & Vermeulen, I. E. (2013). The language of extraversion: Extraverted people talk more abstractly, introverts are more concrete. *Journal of Language and Social Psychology, 32*(2), 191-201.

Bhele, S. G., & Mankar, V. H., (2012). A review paper on face recognition techniques. *International Journal of Advanced Research in Computer Engineering & Technology*, *1*(8), 339-346.

Bond, G. D., & Lee, A. Y. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology*, *19*(3), 313-329.

Bowditch, C., & Hobby, E. (2015). Introduction: Aphra Behn, New questions and contexts. *Women's Writing*, *22*(1), 1-12.

Boyd, R. L. (in press). Psychological text analysis in the digital humanities. In S. Hai-Jew (Ed.), *Data analytics in the digital humanities*. New York City, NY, US: Springer Science.3

Boyd, R. L. (under review). Mental profile mapping: A single-candidate authorship attribution analysis method. Submitted and under review.

Boyd, R. L. (2014a). MEH: Meaning Extraction Helper (Version 1.0.6) [Software]. Available from http://meh.ryanb.cc

Boyd, R. L. (2014b). RIOT Scan: Recursive Inspection of Text Scanner (Version 1.8.3) [Software]. Available from http://riot.ryanb.cc

Boyd, R. L. (2014c). Ye Olde Token Converter (Version 0.2.5) [Software]. Available from http://riot.ryanb.cc/other-tools

Boyd, R. L., & Pennebaker, J. W. (2015a). A way with words: Using language for psychological science in the modern era. In C. Dimofte, C. Haugtvedt, & R.

Yalch (Eds.), *Consumer psychology in a social media world* (pp. 222-236). New York City, NY, US: Routledge Publishers.

Boyd, R. L., & Pennebaker, J. W. (2015b). Did Shakespeare write Double Falsehood? Identifying an individual's psychological signature with text analysis. Psychological Science, 26(5), 570-582.

Boyd, R. L., Wilson, S. R., Pennebaker, J. W., Kosinski, M., Stillwell, D. J., & Mihalcea, R. (2015). Values in words: Using language to evaluate and understand personal values. *Proceedings of the Ninth International AAAI Conference on Web and Social Media*, 31-40.

Brinegar, C. S. (1963). Mark Twain and the Quintus Curtius Snodgrass letters: A statistical test of authorship. *Journal of the American Statistical Association, 58*, 85-96.

Bromby, M. C. (2011). Juries and their understanding of forensic science: Are jurors equipped?. *The International Journal of Science in Society, 2*(2), 247-256.

Bubeck, M., and Bilsky, W. (2004). Value structure at an early age. *Swiss Journal of Psychology, 63*, 31–41

Burke, P. A., & Dollinger, S. J. (2005). A picture's worth a thousand words: Language use in the autophotographic essay. *Personality and Social Psychology Bulletin, 31*(4), 536-548.

Caprara, G. V., Schwartz, S., Capanna, C., Vecchione, M., & Barbaranelli, C. (2006). Personality and politics: Values, traits, and political choice. *Political Psychology, 27*, 1-28.

Carnegie, D., & Taylor, G. (2012). *The quest for Cardenio: Shakespeare, Fletcher, Certantes, and the lost play*. Published Online by Oxford Scholarship Online: September, 2012.

Cassidy, J., Sherman, L. J., & Jones, J. D. (2012). What's in a word? Linguistic characteristics of Adult Attachment Interviews. *Attachment & Human Development, 14*(1), 11-32.

Colman, A. M., Walley, M., & Sluckin, W. (2011). Preferences for common words, uncommon words and non-words by children and young adults. *British Journal of Psychology, 66*(4), 481-486.

Chang, F., Guo, C., Lin, X., & Lu, C. (2010). Tree decomposition for large-scale SVM problems. *Journal of Machine Learning Research, 11*, 2935-2972.

Chung, C.K., & Pennebaker, J.W. (2007). The psychological functions of function words. In K. Fiedler (Ed.), *Social communication* (pp. 343-359). New York: Psychology Press.

Chung, C. K., & Pennebaker, J. W. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality, 42*, 96-132.

Cieciuch, J., Schwartz, S. H., & Vecchione, M. (2013). Applying the refined values theory to past data: What can researchers gain? *Journal of Cross-Cultural Psychology, 44*(8), 1215-1234.

Clark, A., & Aubrey, J., (1898). *"Brief Lives", chiefly of contemporaries, set down by John Aubrey, between the years of 1669 & 1696*. Oxford, MA: At The Clarendon Press.

Corbett, C. (1744). *A catalogue of the library of Lewis Theobald, Esq. deceas'd: among which are many of the classicks, poets and historians, of the best editions*. Eighteenth Century Collections Online. Gale.

Craig, H. (1999). Authorial attribution and computational stylistics: If you can tell authors apart, have you learned anything about them? *Literary and Linguistic Computing 14*(1), 103-113.

Craig, H., & Kinney, A. (2009). *Shakespeare, computers, and the mystery of authorship*. Cambridge: Cambridge University Press.

Curram, S. P., & Mingers, J. (1994). Neural networks, decision tree induction and discriminant analysis: An empirical comparison. *Journal of the Operational Research Society, 45*(4), 440-450.

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. In *Annual Proceedings of the 2013 AAAI Conference on Web and Social Media (ICWSM).*

Depressed. (n.d.). Retrieved June 30, 2014, from http://www.merriam-webster.com/dictionary/depressed

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2014). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114-126.

Dolan, F. E. (2013). *True relations: Reading, literature, and evidence in Seventeenth-Century England.* Philadelphia, PA: University of Pennsylvania Press.

Dominik, M. (1991). *William Shakespeare and the birth of Merlin*. Beaverton, OR: Alioth Press.

Evans, Melanie (2016). By the Queen: Collaborative authorship in scribal correspondence of Queen Elizabeth I (chapter 3). In Daybell, J., & Gordon, A. (Eds.). *Women and Epistolary Agency in Early Modern Culture*, 1450–1690. Routledge.

Dell EMC (2012). *New digital universe study reveals big data gap: Less than 1% of world's data is analyzed; Less than 2% is protected* [Press release]. Retrieved from http://www.emc.com/about/news/press/2012/20121211-01.htm

Fetterman, A. K., Boyd, R. L., & Robinson, M. D. (2015). Power versus affiliation in political ideology: Robust linguist evidence for distinct motivation-related signatures. *Personality and Social Psychology Bulletin, 41*(9), 1195-1206.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*(2), 179-188.

Foklenflik, R. (2012). "Shakespearesque": The Arden Double Falsehood. *Huntington Library Quarterly, 75*(1), 131-143.

Foster, D. W. (1996). Primary culprit: An analysis of a novel of politics. *New York, 29,* 50-57.

Freehafer, J. (1969). Cardenio, by Shakespeare and Fletcher. *PMLA, 84,* 501-513.

Fucks, W. (1952). On the mathematical analysis of style. *Biometrika, 39*, 122-129.

Funder, D. C. (2015). *The personality puzzle (7th edition)*. New York, NY: W. W. Norton & Company, Inc.

Goodwin, C. J., & Goodwin, K. A. (2013). *Research in psychology: Methods and design, 7th edition*. Hoboken, NJ, US: Wiley & Sons Inc.

Graham, L. T., & Sandy, C. J., & Gosling, S. D. (2011). Manifestations of individual differences in physical and virtual environments. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *Handbook of Individual Differences* (pp. 773-800). Oxford: Wiley-Blackwell.

Guastella, A. J., & Dadds, M. R. (2006). Cognitive-behavioral models of emotional writing: A validation study. *Cognitive Therapy and Research, 30*(3), 397-414.

Grieve, J. W. (2007). Quantitative authorship attribution: An evaluation of techniques. *Literary and Linguistic Computing 22*(3), 251-270.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations, 11*(1), 10-18.

Harari, G. M., Lane, N., Wang, R., Crosier, B., Campbell, A. T., & Gosling, S. D. (in press). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*.

Harris, Z. (1954). Distributional structure. *Word, 10*, 146-162.

Hartley, J., Pennebaker, J.W., & Fox, C. (2003). Using new technology to assess the academic writing styles of male and female pairs and individuals. *Journal of Technical Writing and Communication, 33*, 243-261.

Holmes, D. I. (1994). Authorship attribution. *Computers and the Humanities 28*(2), 87-106.

Horowitz, L. M., Wilson, K. R., Turan, B., Zolotsev, P., Constantino, M. J., & Henderson, L. (2006). How interpersonal motives clarify the meaning of interpersonal behavior: A revised circumplex model. *Personality and Social Psychology Review, 10*(1), 67-86.

Hughes, D. (2001). *The theatre of Aphra Behn*. New York, NY: Palgrave Publishers Ltd.

Ireland, M.E., & Pennebaker, J.W. (2010). Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology, 99*, 549-571.

Johnson, E. (1996). *Lexical change and variation in the Southeastern United States: 1930-1990.* Tuscaloosa, AL: University of Alabama Press.

Jones, R. F. (1919). *Lewis Theobald: His contribution to English scholarship with some unpublished letters*. New York: Columbia University Press.

Juola, Patrick. *Authorship attribution*. Foundations and Trends in Information Retrieval 1.3 (2006): 233-334.

Kacmarcik, G., & Gamon, M. (2006). Obfuscating document stylometry to preserve author anonymity. *Proceedings of ACL, 2006*.

Klammer, T. P., Schulz, M. R., &Volpe, A.D. (2012). *Analyzing English grammar, 7th ed*. New York: Longman.

Kolb, J., & Kolb, J. (2013). *The big data revolution*. CreateSpace Independent Publishing Platform.

Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art, and perspective. *Artificial Intelligence in Medicine, 23*(1), 89-109.

Koppel, M., Schler, J., & Argamon, S. (2008). Computational methods in authorship attribution. *Journal of the American Society for Information Science and Technology, 60*(1), 9–26.

Koppel, M., Schler, J., & Argamon, S. (2009). Computational methods in authorship attribution. *Journal of the American Society for Information Science and Technology, 60*(1), 9-26.

Koppel, M., Schler, J., & Argamon, S. (2013). Authorship attribution: What's easy and what's hard?. *Journal of Law and Policy, 21*(2), 317-331.

Koppel, M., & Winter, Y. (2014). Determining if two documents are by the same author. *Journal of the Association for Information Science and Technology, 65*(1), 178-187.

Korte, B. (2015). Aphra Behn's *The Widow Ranter*: theatrical heroics in a strange new world. *Anglia Journal of English Philology, 133*(3), 435-451.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academcy of Sciences, 110*(15), 5802-5805.

Kramer, A. D. I., & Chung, C. K. (2011). Dimensions of self-expression in Facebook status updates. *Proceedings of the Fifth International AAAI Conference on Web and Social Media*, 169-176.

Kristiansen, C. M, & Zanna, M. P. (1988). Justifying attitudes by appealing to values: A functional perspective. *British Journal of Social Psychology, 27*(3), 247-256.

Krippendorf, K. (2013). *Content analysis: An introduction to its methodology*. US: Sage Publications.

Kukowski, S. (1990). The hand of John Fletcher in Double Falsehood. *Shakespeare Survey, 43*, 81-89.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Anlysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211-240.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., & Christakis, N. (2009). Computational social science. *Science, 323*(2915), 721-723.

Leigh, L. (2011). "'Tis no such killing matter": Rape in Fletcher and Shakespeare's Cardenio

Lerner, K. L., & Lerner, B. W. (2005). Hitler diaries. *World of Forensic Science, 1*, 348-350.

Lepley, R. (1957). *The language of value*. New York: Columbia University Press.

Lowe, R. D., Heim, D., Chung, C. K., Duffy, J., Davies, J. & Pennebaker, J. W. (2013). In verbis venum? Relating themes in an open-ended writing task to alcohol behaviors. *Appetite, 68*, 8-13.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences in India, 2*(1), 49-55.

Martindale, C. (1975). *Romantic progression: The psychology of literary history*. Washington, D.C.: Hemisphere.

Martindale, C., & McKenzie, D. (1995). On the utility of content analysis in authorship attribution: The Federalist Papers. *Computers and the Humanities, 29*, 259-270.

McAdams, D. P. (2006). *The redemptive self: Stories Americans live by*. New York: Oxford University Press.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani & K. Weinberger (Eds.), *Advances in neural information processing systems, 26th annual proceedings* (pp. 3111-3119). Curran Associates, Inc.

Miller, G. A. (1995). *The science of words.* New York: Scientific American Library.

Morton, A. Q. (1978). *Literary detection: How to prove authorship and fraud in literature and documents*. New York: Scribner's.

Mosteller, F., & Wallace, D. L. (1964). *Inference and disputed authorship: The Federalist.* Reading: MA: Addison-Wesley, 1964.

Nguyen, K., & Ock, C. (2013). Word sense disambiguation as a traveling salesman problem. *Artificial Intelligence Review, 40*, 405-427.

O'Donnell, M. A. (2004). Chronology and Aphra Behn: The documentary record. In Hughes, D., & Todd, J. (Eds.), *The Cambridge companion to Aphra Behn* (preface, 1-11). Cambridge, UK: Cambridge University Press.

Pennebaker, J. W. (2004). *Writing to heal: A guided journal for recovering from trauma and emotional upheaval*. Oakland, CA: New Harbinger Press.

Pennebaker, J. W. (2007). Current issues and new directions in Psychology and Health: Listening to what people say--the value of narrative and computational linguistics in health psychology. *Psychology & Health, 22*(6), 631-635.

Pennebaker, J. W. (2011). *The secret life of pronouns*: *What our words say about us*. New York: Bloomsbury.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates (www.liwc.net).

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic Inquiry and Word Count (LIWC2007): A computerized text analysis program. Austin, Texas: LIWC.net

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014) When small words foretell academic success: The case of college admissions essays. *PLOS ONE 9*(12): e115844.

Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A., & Booth, R.J. (2007). *The development and psychometric properties of LIWC2007*. [Software manual]. Austin, TX: LIWC.net

Pennebaker, J.W., & Ireland, M.E. (2011). Using literature to understand authors: The case for computerized text analysis. *Scientific Study of Literature, 1*, 34-48.

Pennebaker, J. W., & King, L. A (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*(6), 1296-1312.

Pennebaker, J.W., Mehl, M.R., & Niederhoffer, K.G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547-577.

Petrie, K.J., Pennebaker, J.W., & Sivertsen, B. (2008). Things we said today: A linguistic analysis of the Beatles. *Psychology of Aesthetics, Creativity, and the Arts, 2*, 197-202.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Developmental Psychopathology, 17*(3), 715-734.

Potter, L. (2012). *The life of William Shakespeare: A critical biography*. Wiley-Blackwell Publishing.

Promberger, M., & Baron, J. (2006). Do patients trust computers? *Journal of Behavioral Decision Making, 19*, 455-468.

Ramirez-Esparza, N., Chung, C. K., Kacewicz, E., & Pennebaker, J. W. (2008). The psychology of word use in depression forums in English and in Spanish: Testing

two text analytic approaches. *Proceedings of the 2008 International Conference on Weblogs and Social Media*, pp.102-108.

Robinson, M. D., Boyd, R. L., & Liu, T. (2013). Understanding personality and predicting outcomes: The utility of cognitive-behavioral probes of approach and avoidance motivation. *Emotion Review, 5*(3), 1-5.

Rockeach, M. (1968). *Beliefs, attitudes, and values, Vol. 34*. San Francisco, CA: Jossey-Bass.

Rockeach, M. (1970). *Beliefs, attitudes, and values, Vol. 70*. San Francisco, CA: Jossey-Bass.

Rogers, P. (Ed., 2004). *The Alexander Pope encyclopedia.* Westport, CT: Greenwood Press

Sagi, E., & Dehghani, M. (2014). Measuring moral rhetoric in text. *Social Science Computer Review, 32*(2), 132-144.

Savova, G. K., Coden, A. R., Sominsky, I. L., Johnson, R., Ogren, P. V., de Groen, P. C., & Chute, C. G. (2008). Word sense disambiguation across two domains: Biomedical literature and clinical notes. *Journal of Biomedical Informatics, 41*(6), 1088-1100.

Schinka, J. A., Velicer, W. F., & Weiner, I. B. (2003). *Handbook of psychology: Research methods in psychology (Volume 2)*. Hoboken, New Jersey: John Wiley & Sons.

Schwartz, H. A., Eichstaedt, J., Blanco, E., Dziurzynski, L., Kern, M. L., Ramones, S., Seligman, M., … et al. (2013a). Choosing the right words: Characterizing and

reducing error of the word count approach. In *Proceedings of the Second Joint Conference on Lexical and Computational Semantics (*SEM)*. Atlanta, GA, USA.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., … et al. (2013b). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS One*. Published online: September 25, 2013.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology, 25*, 1-65.

Schwartz, S. H. (1994). Beyond individualism/collectivism: New cultural dimensions of values. *Cross-cultural Research and Methodology Series, 18*, 85-119.

Schwartz, S. H. (2004). Mapping and interpreting cultural differences around the world. In Vinken, H., Soeters, J., and Ester, P. (Eds.), *Comparing cultures: Dimensions of culture in a comparative perspective* (pp. 43-73). Leiden, Netherlands: Brill.

Schwartz, S. H. (2009). *Proper use of the Schwartz Value Survey*, *version 14*. Technical report.

Schwartz, S. H., & Huismans, S. (1995). Value priorities and religiosity in four Western religions. *Social Psychology Quartery*, *58*(2), 88-107.

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fisher, R., Beierlein, C., Ramos, A., … & Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, *103*(4), 663-688.

Seary, P. (1990). *Lewis Theobald and the editing of Shakespeare*. New York: Oxford University Press.

Shaffer, D. R. (2000). Physical illness and depression in older adults. In Williamson, G. M., Shaffer, D. R., & Parmelee, P. A. (Eds.). *Physical illness and depression in older adults: A handbook of theory, research, and practice*. Boston, MA: Springer US.

Shakespeare, W. (2010). B. Hammond (Ed.), *Double falsehood* (3rd Edition). London: A & C Black Publishers Ltd.

Shakespeare, W., Irving, H., Marshall, F. A. (Ed.), & Dowden, E (Ed.). (1890). *The works of William Shakespeare*. New York: Scribner and Wellford.

Solan, L. M. (2013). Intuition versus algorithm: The case of forensic authorship attribution. *Brooklyn Journal of Law and Policy, 21*(551), 551-576.

Spencer, J. (2000). *Aphra Behn's afterlife*. New York, NY: Oxford University Press.

Stern, T. (2011). "The forgery of some modern author"?: Theobald's Shakespeare and Cardenio's Double Falsehood. *Shakespeare Quarterly, 62*, 555-593.

Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine, 63*, 517-522.

Strappavara, C., & Mihalcea, R. (2008). Learning to identify emotions in text. *Proceedings of the 2008 ACM symposium on applied computing*.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24-54.

Todd, J. M. (1998). *The critical fortunes of Aphra Behn*. Columbia, SC: Camden House.

Tsai, C., Lai, C., Chao, H., & Vasilakos, A. V. (2015). Big data analytics: A survey. *Journal of Big Data, 2*(21). doi:10.1186/s40537-015-0030-3

Ule, L. (1982). Recent progress in computer methods of authorship determination. *Association for Literary and Linguistic Computing Bulletin, 10,* 73-89.

van Halteren, H., Baayen, R. H., Tweedie, F., Haverkort, M., & Neijt, A. (2005). New machine learning methods demonstrate the existence of a human stylome. *Journal of Quantitative Linguistics 12*(1), 65–77.

Vickers, B. (2004). *Shakespeare, co-author*. New York, NY: Oxford University Press, Inc.

Vickers, B. (2011). Shakespeare and authorship studies in the twenty-first century. *Shakespeare Quarterly, 62*, 106-142.

Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision, 57*(2), 137-154.

Wellman, F. L. (1936). *The art of cross-examination (4th edition)*. New York: MacMillan.

Wolf, M., Horn, A., Mehl, M., Haug, S., Pennebaker, J. W., & Kordy, H. (2008). Computer-aided quantitative text analysis: Equivalence and reliability of the German adaptation of the Linguistic Inquiry and Word Count. *Diagnostica, 54*(2), 85-98.

Witten, I. H., Frank, E., & Hall, M. A. (2011). Data mining: *Practical machine learning tools and techniques* (3rd edition). Elsevier Inc.: Burlington, MA.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality, 44*(3), 363-373.

Yule, G. U. (1944). On sentence-length as a statistical characteristic of style in prose, with application to two cases of disputed authorship. *Biometrika*, *30*, 363-390.

Zhang, Y., Chen, W., Wang, D., & Yang, Q. (2011). User click modeling for understanding and predicting search behavior. *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining*, 1388-1396.